Piedmont
College Library

DEMOREST, GEORGIA

16389

LB 1131 .S64
Smith, Henry Lester
Tests and measurements

370.15 Sm5

Digitized by the Internet Archive
in 2026

https://archive.org/details/testsmeasurement0000henr

Tests and Measurements

Tests and Measurements

by

HENRY LESTER SMITH

DEAN OF THE SCHOOL OF EDUCATION
INDIANA UNIVERSITY

and

WENDELL WILLIAM WRIGHT

ASSOCIATE PROFESSOR OF EDUCATION
INDIANA UNIVERSITY

SILVER, BURDETT AND COMPANY
NEW YORK NEWARK BOSTON CHICAGO SAN FRANCISCO

COPYRIGHT, 1928,
BY SILVER, BURDETT AND COMPANY

Printed in the United States of America

PREFACE

Tests and Measurements have been treated in most books as a separate and distinct entity in the field of education. The plan of this book is to bring together the objectives of the teaching of a particular subject and the measurement of the results of this teaching in so far as these results can be measured in the achievement of pupils. Testing, when approached from this point of view, becomes an integral part of the whole educational procedure.

A second contribution made by *Tests and Measurements* is the help that it gives the teacher in the use of standard tests to diagnose pupil difficulties and to suggest remedial measures.

In a recent bibliography approximately eight hundred standard tests are described. Many of these tests are crude and poorly adapted to use in the classroom. Some method must be set up for evaluating tests by the teacher. This book not only attempts to call special attention to various good standard tests, but also to set up a method of evaluating tests.

The wide use of informal objective or new-type examinations makes it necessary that teachers, supervisors, and administrators understand the guiding principles and the technique underlying their development. *Tests and Measurements* elucidates the principles underlying the construction and use of objective or new-type examinations. It further suggests methods of construction and shows by concrete illustrations how they are applied. Since intelligence tests, prognosis tests, and practice tests have come into common use, it is not only important that teachers, supervisors, and administrators should know about them, but necessary that they should have a basis for judging them critically. *Tests and Measurements* attempts to bring all these factors forcibly to the reader.

The authors have drawn freely from many sources. They hereby acknowledge many courtesies extended by authors and publishers of standard tests as well as by teachers and superintendents who have used the tests.

H. L. SMITH,
W. W. WRIGHT.

Indiana University.

CONTENTS

CHAPTER		PAGE
I	Introduction	1
II	Measuring and Reporting Achievement	11
III	Criteria for Judging Standardized Tests	32
IV	Interpreting and Reporting Scores	43
V	Arithmetic	68
VI	Measurement in Secondary Mathematics	109
VII	Language	132
VIII	Spelling and Handwriting	170
IX	Reading	196
X	History and Civics	235
XI	Geography	268
XII	Music and Art	282
XIII	Science	302
XIV	Foreign Language	336
XV	Industrial Arts and Vocational Subjects	365
XVI	Commercial Subjects	385
XVII	Health and Physical Education	400
XVIII	Practice Tests	422
XIX	Intelligence Tests	440
XX	Prognosis and Special Ability Tests	463
XXI	New-Type Examinations	481
XXII	Construction of Achievement Tests	511
	Index	i

136.500

CHAPTER I

Introduction

Measurement of the effects of teaching and learning has long been attempted on a subjective basis. The teacher has said that a child succeeded or that he failed; that he made his grade or that he did not make it; that the quality of his failing work was very poor, poor, or just below passing; that his passing work was just above passing, fair, good, or excellent.

As early as 1845, according to S. A. Courtis, a committee attempted to measure the efficiency of the Boston schools; and Horace Mann in his comments upon the report prophesied that the future would see tests scientifically prepared, given, scored, and tabulated.

In England, according to E. L. Thorndike, George Fisher, connected with the Hospital School at Greenwich, anticipated as early as 1864, in his "scale book," present-day standards of school measurement. In this scale book were standard specimens with which the material to be measured was to be compared. For instance, Ficher had selected various specimens of handwriting, and had placed them in the scale book. Specimens of handwriting which he wished to evaluate were compared with the standard handwriting specimens in the scale book.

Conditions were not favorable, however, to the spread of this measurement movement anticipated by Horace Mann in the United States and worked out somewhat crudely by Ficher in England. There is no evidence for years afterwards that educators in either country gave serious consideration to the further development of the measurement movement in education along the lines suggested by these pioneers.

In 1894, J. M. Rice had an inspiration that culminated a few months later in the beginning of research in spelling, covering a period of sixteen months. His general plan was published in the *Forum,* December, 1896, under the title "Obstacles to Rational Educational Reform." At the 1897 meeting in Indianapolis of the Department of Superintendence, National Education Association, Rice, conducting a round-table discussion, startled those present by asking, in substance, how any one could tell whether at the end of eight years' training children who had had forty minutes a day in spelling instruction were better spellers than those who had had only ten minutes a day. Vigorous opposition developed against the implications of Rice's question and the testing movement in education lay dormant for more than another decade.

While Rice's articles were appearing in the *Forum,* however, a student in Columbia University, E. L. Thorndike, became impressed with the suggestions raised by Rice's work.

One of the first results of Thorndike's thinking along the line of measurement was the material he prepared in 1907 for the United States Bureau of Education, entitled "The Elimination of Pupils from School." Age-grade tables as well as statistics on elimination were used in this study as a measure of a pupil's progress. A further study in this same field was made in 1911 by G. D. Strayer under the title of "Age and Grade Census of Schools and Colleges; A Study of Retardation and Elimination." [1] In 1913 Leonard P. Ayres brought out his classic study entitled "Laggards in Our Schools; A Study of Retardation and Elimination in City School Systems." [2]

In the meantime measurement in education was branching out in another direction. In 1911 appeared the preliminary report of the National Education Association Committee on Uniform Records and Reports. The work of this Committee resulted in a very rapid standardization of recording and reporting pro-

[1] United States Bureau of Education Bulletin, 1911, No. 5.
[2] Russell Sage Foundation Publication.

cedures in the field of pupil accounting. The Committee made its final report in 1912. By 1919 a Score Card for City School Buildings had been prepared by G. D. Strayer and published by the National Society for the Study of Education.

Thorndike's influence in the field of measurement was not limited to a study of elimination and of age and grade progress. He had likewise formulated the principles and ideals of testing, and in December, 1909, had presented his Handwriting Scale before the Boston meeting of the American Association for the Advancement of Science. This scale was published the following year and represented the first accomplishment in the field of standardized scales for the measurement of classroom achievement of pupils in the public schools of the country. S. A. Courtis soon followed with his arithmetic tests. Numerous survey tests have appeared since.

As early as 1912 Burke, of the San Francisco State Normal, and Courtis, of Detroit, had done pioneer work in the development of individual lesson sheets which served as a standard procedure for remedial work following the revelations of survey tests. The development of individual remedial work was based upon the findings from another type of test that had been evolving parallel to the development of achievement tests. This other type is the intelligence test. This type of test first appeared in Paris in 1908, having been worked out by Binet and Simon. In the United States Goddard, Yerkes, Terman, Thorndike, Whipple, and others rapidly pushed forward further inventions in this field.

Since 1914, opposition to scientific measurement of schoolroom product has died down. In the 1913 school survey of New York City, standardized achievement tests were used. Similar tests were used in the state-wide survey by the Virginia Education Commission in 1919. Intelligence tests were also used. No comprehensive school survey since 1919 has failed to use achievement tests.

Educational measurement, in so far as it is treated in this book, is based upon certain assumptions, four of which may be stated as principles. It is not our purpose to prove the validity of these statements of principles but rather to interpret testing in the light of them. Some of these statements of principles, to be sure, are not proved facts. However, they are working theories, rather generally accepted as basic to the testing movement. These principles are[3]:

1. "Whatever exists at all exists in some amount."[4] The amount may be measured in quality or quantity or both. The measurement made is of the product and of the change as a reflection of the inner condition of the individual measured.

This principle may not be universally accepted in theory, but it is universally accepted in practice. There is no single quality that can be described by words that is not measured either in minute detailed units or by gross comparison. From the measure of the power of the right hand's grip to the degree of honesty of Preacher X., measurements are made, either by instruments perfected for the purpose of accurate measurement of the former or by the opinions of the members of the congregation concerning the latter. Many of the factors that exist in the human being are not directly measurable as to amount, although the change effected by their presence is directly measurable. Wherever change is effected, this change is measurable even though the scientific instruments for its measurement are not yet perfected. The thermometer was not invented until 1782. Yet one can hardly believe that changes in temperature had not been noticed before that scientific instrument for measuring them was perfected. Not all the qualities of human beings, and especially of children, can be accurately and scientifically measured with the means of measurement now at hand. In many cases the change effected in the pupil by a teacher's work with him cannot as yet be scientifically measured. There

[3] Cf. McCall's *How to Measure in Education* (Macmillan), Chap. I.
[4] *17th Yearbook*, National Society for the Study of Education, p. 16.

is much reason to believe, however, that many if not all of the factors of human knowledge, qualities, attitudes, and skills can be measured.

It should be clearly understood that the thesis, "whatever exists at all exists in some amount and can be measured," should not in any sense be interpreted as meaning that scientific instruments are now available to perform such functions. The science of measurement is very young and can only incompletely perform any service that it attempts. This fact in no way affects the validity of the foregoing thesis as a working principle. It does, however, serve as a challenge to the efforts of those who attempt to develop measuring instruments, and it serves as a stimulus to those who are attempting to measure the more subtle qualities of human behavior through the products of their reactions.

2. Scientific measurement in education is closely related to educational progress. Effectiveness in the field of medicine would be decidedly reduced without the use of the thermometer. It would likewise be difficult to know the results of dietetic feeding except as it is possible to measure accurately change in weight. In the same manner, the effectiveness of the teaching of spelling is decidedly reduced if we pay no attention to measuring how well children can spell before and after they have been taught spelling. Of course the methods of measurement in education are not in any sense confined to tests. Tests are only one of the measures being developed to further the progress of education. The fact that many phases of education are not as yet scientifically measurable by delicately perfected instruments is no excuse for not using the crude measures that are available.

3. Scientific measurement in education is conditioned by the objectives of education as a whole and of school subjects specifically. "Ask me another one" or "Ask me one I can't answer" may be a good slogan for a question and answer book, but its relation to that phase of scientific measurement in education

called standard tests is nihil. The objectives of education whose selection is much influenced by the social utility point of view in education would not justify the inclusion in our arithmetic for elementary schools of such an item as cube root. It would, then, be foolish to include in a standard arithmetic test for elementary school pupils a problem in cube root. If the trend of emphasis in education is away from factual knowledge in the social sciences and toward interpretation of facts, solving of social problems, and development of the emotions, then testing in social sciences should follow in these newer directions instead of lingering in the field of memory of factual knowledge.

The greatest single deviation from principle 3 is found in the practice of using tests that test only a part of the objectives of any subject and of interpreting those results as the measure of the total achievement in that subject. The ability to locate places on a map is somewhat important and basic to the understanding of geography. A measure of ability to do this, however, is no assured measure of how much of a sympathetic understanding the pupil may have of the peoples of other lands. Until it is proved that these two factors are so closely correlated that a measure of one may be a measure of the other, the test of ability to locate places on a map cannot be called a geography test *per se*. It is geography place testing and nothing more.

Another deviation from this principle is found in the practice of measuring achievement in the traditional three r's and assuming that this is a complete measure of the work of the public schools. In order to get a measure of the total progress of education toward the desired goal, every aim of education should be considered, and the progress toward the achievement of each of these aims should be measured. For example, in most of the recent summaries of the aims of education, health development and maintenance are included. If it is desired, therefore, to get a complete measure of the achievement of the school, it will be necessary to make our measurements comprehensive enough to include progress made in the development

and maintenance of health as well as in the traditional three r's.

4. Educational measurement is justified only by the resulting improvement in learning on the part of the child. Measurement for measurement's sake—that is, measuring to develop more accurate and scientific measure—is only for a few scientists in the field of work. Their contribution is that of the pure scientist. The contribution that is thus made may be far-reaching and exceedingly important. In the ordinary school system, however, there is no place for testing for testing's sake. Testing merely to get the results all carefully arrived at and filed away, without following the results up to see that they function in better administering, supervising, and teaching so that the child can do more effective learning, is the greatest of all the sins of standardized testing. "I have given the children in my school standard tests several times, and it has not done them a bit of good," was the complaint of one superintendent of schools in a summer school class. Superintendents' offices and the research laboratories of schools of education are sometimes filled with tests and the results of testing carefully arrived at and never used, at least to the extent of their possibilities. It should be understood that 90 per cent of the value of testing depends on what is done after the tests have been selected, given, and scored, provided this phase of the work has been well done. Correct interpretation of results and the remedial administration, supervision, or teaching that follow, which will aid the child to learn more effectively, are the biggest values in testing.

BIBLIOGRAPHY

Buckingham, B. R. "Standards in Education," *Journal of Educational Research,* Vol. 10: pp. 1-18, June, 1924.

Cardozo, F. L. "Tests and Measurements in Public Schools," *School and Society,* Vol. 20: pp. 797-798, December 20, 1924.

Douglas, H. R. "Some Uses and Limitations of the Standard Educational Test," *Educational Administration & Supervision,* Vol. 5: pp. 474-490, December, 1919.

Greenwood, E. "Grading Human Beings: Achievement Tests, *Independent,* Vol. 115: pp. 557-558, November 14, 1925.

Haggerty, M. E. "Character of Educational and Scientifi Method," *Journal of Educational Research,* Vol. 13: pp. 233 248, April, 1926.

Henmon, V. A. C. "Some Limitations of Educational Tests, *Journal of Educational Research,* Vol. 7: pp. 185-198, March 1923.

McCall, W. A. *How to Measure in Education,* Chap. 1. Mac millan Company, 1922-23.

Manuel, H. T. "After Measuring, What?" *The Journal of th National Education Association,* Vol. 15: pp. 139-141, May 1926.

Myers, G. C. "Teaching Versus Testing," *American School Boar Journal,* Vol. 72: p. 47, February, 1926.

Neville, Charles E. "Supervision Through Simplified Testing, *Elementary School Journal,* Vol. 25: pp. 696-99, March, 1925

Thorndike, E. L. "The Nature, Purposes and General Methods o Measurements of Educational Products," *National Society fo the Study of Education,* 17th Yearbook, Part II, Chapter 2 pp. 17-25, 1918.

Toops, H. A. "What Are We Failing to Measure in Education?" *Journal of Educational Research,* Vol. 13: pp. 118-128, Feb ruary, 1926.

Toops, H. A. "A Prophecy of the Future of Tests and Measure ments in Education," *Thirteenth Annual Schoolmen's Week Proceedings,* University of Pennsylvania, March, 1926.

The following is a general bibliography that will be found applicable to several chapters in the book.

Brooks, Samuel S. *Improving Schools by Standardized Tests* Houghton Mifflin Company, 1922.

Doherty and MacLatchy. "Bibliography of Educational and Psy chological Tests and Measurements," *Bulletin No. 55, 1923* Bureau of Education, Department of the Interior, Washing ton, D. C.

Freeman, Frank N. *Mental Tests.* Houghton Mifflin Company 1926.

Greene and Jorgensen. *The Use and Interpretation of Educationa Tests.* Longmans, Green and Company, 1929.

Gregory, Chester Arthur. *Fundamentals of Educational Measure ment.* D. Appleton and Company, 1924.

Hines, Harlan C. *A Guide to Educational Measurements.* Houghton Mifflin Company, 1923.

Kelley, Truman Lee. *Interpretation of Educational Measurements.* World Book Company, 1927.

Lincoln, Edward A. *Beginnings in Educational Measurement* (Revised). J. B. Lippincott, 1927.

Monroe, DeVoss, and Kelly. *Educational Tests and Measurements.* Houghton Mifflin Company, 1924.

Monroe and Asher. "A Bibliography of Bibliographies." *Bulletin No. 36, Bureau of Educational Research,* University of Illinois, 1927.

McCall, William A. *How to Measure in Education.* The Macmillan Company, 1923.

Odell, Charles W. "A Glossary of Three Hundred Terms Used in Educational Measurement and Research." *Bulletin No. 40, Bureau of Educational Research,* University of Illinois, 1928.

Odell, Charles W. *Educational Statistics.* The Century Co., 1925.

Odell, Charles W. "Educational Tests for Use in Elementary Schools, Second Revision." University of Illinois Bulletin, Vol. XXIV, *Educational Research Circular No. 49.*

Odell, Charles W. "Educational Tests for Use in High Schools, Second Revision," University of Illinois Bulletin, Vol. XXIV, *Educational Research Circular No. 48.*

Otis, Arthur S. *Statistical Method in Educational Measurement.* World Book Co., 1925.

Pintner, Rudolf. *Intelligence Testing.* Henry Holt and Company, 1925.

Pressey, Sidney L. and Pressey, Luella Cole. *Introduction to the Use of Standard Tests.* World Book Company, 1923.

Ruch, G. M. and Stoddard, George D. *Tests and Measurements in High School Instruction.* World Book Company, 1927.

Rural School Survey of New York State. "Educational Achievement," M. E. Haggerty, 1923.

Russell, Charles. *Classroom Tests.* Ginn and Company, 1926.

School Surveys. The reports of the school surveys of Tampa, Florida; Watertown, New York; Port Arthur, Texas, made by the Division of Field Studies, Institute of Educational Research, Teachers College, Columbia University, New York City. Director, George D. Strayer, Assistant Director, N. L. Engelhardt.

Smith, Henry Lester and Wright, Wendell W. *Bibliography of Educational Measurements* (Second Revision). The Bureau of Coöperative Research, Indiana University, November, 1927.

Symonds, Percival M. *Measurement in Secondary Education.* The Macmillan Company, 1927.

Trabue, Marion Rex. *Measuring Results in Education.* American Book Company, 1924.

Van Wagenen, M. J. *Educational Diagnosis and the Measurement of School Achievement.* The Macmillan Company, 1926.

Wood, Ben D. *Measurement in Higher Education.* World Book Company, 1923.

CHAPTER II

Measuring and Reporting Achievement

School marks are an important, useful, and integral part of education. At almost every level of education from the first grade through the university, marks record the achievement of pupils and students. These marks, which are reported to parents and children, are generally taken seriously. School records are thought to be quite incomplete if they do not show a statement of the child's promotion from grade to grade and also the relative proficiency of his work as indicated by marks received in various subjects. School marks have much to do with the educational and vocational future of the child. These marks often form the basis for the child's own judgment of his ability to master the subjects he has studied and are also the basis for guidance by parents and teachers. Institutions of higher learning often in large part determine the question of entrance by marks made in high school and grant scholarships to those whose marks signify unusual achievement.

The purposes of school marks may be roughly stated as (1) administrative, (2) educational, and (3) social. In promotion as well as in certification or graduation of pupils, the administrative work of the school is very much concerned with school marks, as the various units of the school find difficulty in functioning properly unless there is a consistent system of marking used as a basis for promotion. Study of an age-grade table of almost any school system will reveal congestion in the fourth, fifth, and sixth grades as a result of over-age piling up in these grades. A study of a grade-to-grade promotion of pupils will show likewise that much of the condition of over-ageness in these grades is due to failure to make yearly progress

from grade to grade. A study of the grades in which children have failed to be promoted reveals the incidence of retardation. A study of the causes of retardation will reveal very different bases for determining the passing mark. The "passing mark" is a critical point on a scale of marks. If a pupil falls below this point, he is supposed to be unable to do satisfactorily the work of the succeeding grade. Administrative difficulties arise because teachers are often very uncertain as to the achievement and ability of individual pupils, and most pupils and parents are not satisfied with a failing mark that does not designate the deficiency. The foregoing conditions have led to two worthwhile educational practices. One of them, provisional promotion, permits a pupil whose achievements and ability do not clearly entitle him to promotion to attempt the work of the next grade or next subject. If he shows that he is capable of doing the work, then he may continue. This provisional promotion is sometimes allowed only if the child will attend summer school. Another result of the attempt to overcome the difficulties and dissatisfaction resulting from non-promotion is the practice of making specific statements, in the course of study and to the child, of the minimum essentials for promotion from any given grade or subject. The following statement is an example of this latter practice.

STANDARDS FOR PROMOTION TO GRADE IV[1]

a. Ability to count and write numbers to 1,000.
b. Ability to add with standard speed and accuracy. Pages 53, 54 (Test I), 64, 116, 134, 140.
c. Ability to subtract with accuracy and reasonable speed. Pages 51, 54 (Test II), 64, 116, 135, 190.
d. Ability to multiply accurately and with reasonable speed numbers of three orders with a one-place multiplier. Pages 80, 82, 109, 116, 135 (Test V).

[1] Department of Public Instruction, Indiana, Bulletin No. 479. Course of Study in Arithmetic, 1924.

e. Ability to divide accurately and with reasonable speed numbers of three orders with a one-place divisor without or with a remainder. Pages 97, 99, 102, 109, 116, 136, 138.
f. Ability to apply the processes in problems involving the measures of time, temperature, weight, capacity, and area. Pages 63, 67 (A), 84 (B), 85 (B), 86, 90, 102, 112, 114, 117, 128, 130, 131, 139.
g. Habits of neat, accurate, and rapid work at seats and on the board.
h. Habits of checking computations.
i. Ability to make a simple drawing to scale—as of the schoolroom. Pages 89, 119.

The analysis of the work of a year in Arithmetic into its unit skills, knowledges, abilities, and attitudes, and the statement of the minimum of these units expected of any pupil mark a significant advance over past practice. Such an analysis and statement serve as an aid not only to marking but also to the development of teaching methods. If, for example, it is known that there are 225 higher decade addition combinations whose sums do not exceed 39, and if it is further known that these represent specific skills to be attained, that certain of them are more difficult than others for children to learn, and that these can all be learned by a third-grade pupil, then this analysis of an important field in Arithmetic should result in better teaching and also in better marking of pupil progress.

Because of the administrative use made of school marks in promotion and educational guidance, it is highly necessary that these marks be on a sound, accurate basis. It is not uncommon for teachers and parents to advise a pupil because of his failure in certain lines of work, as indicated by his marks, not to attempt certain other lines of work, and in some cases to deny these opportunities to him altogether. The school mark would be a most helpful guide as to what the special interests and aptitudes of a pupil were if its meaning, consistency, and reliability could be trusted. School marks are perhaps more valid than they are reliable and consistent. They are valid in the sense that they

do measure what the teacher has in mind. That is, these marks indicate a measure of what they are supposed to measure at the time they are given. However, what the teacher has in mind when Mary is given a mark of 84 in reading at one time and what she has in mind when Mary is given a mark of 86 the next time are often not the same thing. This personal variability of the teacher makes a teacher's marking inconsistent and unreliable.

School marks are often used to motivate school work, since pupils certainly work to some extent for marks. This practice is desirable if the marks are consistent with true achievement. To work for school marks is further desirable if the interpretation of them by pupils results in an enlightenment not only concerning the accomplishment reached but also concerning the needs and shortcomings revealed. Pupils as well as parents often desire and frequently require a statement in terms of pupil achievement. In view of the fact that school marks are taken seriously by both pupils and parents and in view of the possibility of making these marks indicate particular difficulties and needs, marks may be utilized to promote learning. The competition of pupils one against another, or especially that of a pupil against his own record, offers a worthy incentive to activity that will be purposeful and valuable insofar as the mark given is a definite and trustworthy measure of specific achievement. The motive power of the mark varies with the type of pupil, i.e., with respect to ability. A characteristic of children much below the average in general ability is that they desire immediate reward or mark of achievement. If the measure of achievement shows appreciable individual gain, it is especially valuable in motivation. This is illustrated by the joy of a pupil who "raises his grade." A graph of the spelling or arithmetic achievements of an individual pupil is often used by the teacher to motivate further the work in these subjects.

Whenever the procedure in marking is such that pupil and teacher working independently reach the same result in scoring

the pupil's work, the conditions are most satisfactory. The desirability of such a procedure is readily accepted in theory, but in actual practice relatively little has been done toward this end. Because of the desirability of such a marking procedure, practice tests in certain subjects have recently been developed. Such tests are now being extensively and satisfactorily used. The pupil can himself score these practice tests and compare his achievement with standard scores. This practice of pupil and teacher independently grading achievement, each with the standard scale of measurement before him, will come to be resorted to more and more as the minimum essentials of school subjects are more clearly and definitely determined and stated.

Marks have a social significance in the life of a pupil. Many colleges and universities send home those pupils who, according to marks, fail to achieve certain standards in a number of subjects. This practice may have a very important bearing on the lives of the small per cent of students affected. Pupils who fail to attain certain standard marks are often prevented from participating in certain activities. Some schools, in more recent years, have been giving certain privileges to pupils who make higher marks. They have, moreover, been calling upon these pupils for school leadership and extra participation in the life of the school. In one school, appointment to the staff of the school paper represents a special privilege, and only pupils with unusually high marks are in any way considered for such places. A record of consistently high marks over a period of years in school is an excellent sign of more than average ability. Any pupil whose experience with different teachers in many grades and with various types of school work is such that he is consistently marked far above the average for his group must have abilities not found in the average person. This fact, however, does not prove the converse of this statement. Many pupils whose marks are not consistently high possess more than average ability in particular lines of work. Many pupils also, whose school marks are below the average, possess much ability in

lines of endeavor outside the school. The fact that the employers looking for persons of superior ability do depend upon school marks for guidance makes it necessary that the marks be an accurate and consistent measure of achievement or of the ability to achieve.

The greatest source of inaccuracy, inconsistency, and meaninglessness in marks is that teachers use widely different bases and standards. These bases may be achievement in school, improvement in school work, effort, attitude, interest, neatness, habits of work, conduct in class, results of examinations, etc. The fact that the aims of education include good conduct, effort, and right attitude tends to make many teachers feel that the attainment of these ends should be rewarded in the arithmetic mark. The situation is unsatisfactory because we have not only a hybrid measure of achievement but also unknown factors comprising this hybrid measure. Such a measure may be made up of many factors not generally expected or assumed in the measure. This confusion makes for unreliability in school marks, which cannot be prevented until we come to a common understanding of the various factors that should be considered in determining the marks given and of the relative value to be given to each factor. A still more rational and scientific procedure would be to indicate achievement in school subjects by one set of marks and to represent such qualities as conduct and attitude by a separate system of marks. In the general autocracy of many present day school systems, the school subject achievement mark is held tyrannically as a club over the heads of pupils to compel obedience to antiquated and unreasonable rules and regulations of classroom and school system. This practice becomes increasingly prevalent in the upper years of the public school system. The development of school marks to represent achievement in terms of stated essentials analyzed into skills, habits, knowledges, and attitudes that can be understood by teacher and pupil alike to mean the same, will help to eliminate the present chaotic state of school marks.

In their honest endeavor to mark accurately, teachers often encounter difficulty not alone because they have only a vague notion of the desired skills, habits, and attitudes to be attained, but also because they lack standards by which to judge achievement. How legibly should a child in the fifth grade write? To receive a mark of A what per cent of the words of a given list should a pupil in a fourth-grade class spell? Standardization of the school mark in terms of achievement is slowly coming to be accepted as necessary to validation of the school mark. Some attempts have been made along the line of standardization. One of these attempts is well represented by Figure 1, page 18.

The normal-distribution curve, which has been found by scientific investigators to apply to the distribution of a very large number of biological traits, has been accepted rather generally as applying to measures of mental ability and school achievement. The normal-distribution curve is represented by a graph (see Figure 2). This graph shows the application of the normal curve to school marks when five different marks are used. Take for example a school system in which the following marks are given: A (the highest), B, C, D, E (failure). These marks are also interpreted as follows for those who have been used to the percentage system: A = 93 to 100; B = 85 to 92; C = 77 to 84; D = 70 to 76; E = below 70 or failure. If there are 100 children in a grade and they are a rather normal group, how should the marks be distributed? That is, how many should receive A's, how many B's, and how many each of the other marks? If the school achievements distribute themselves according to the normal distribution curve as do biological traits—and it is pretty generally conceded that they do—the number of children receiving A's will be approximately 7, the number of children receiving B's will be approximately 24, the number receiving E's or failing will be approximately 7.[2]

Not always, of course, do marks distribute themselves thus,

[2] Educators vary in their estimates as to the per cent of cases in a normal distribution curve that should receive each grade indicated.

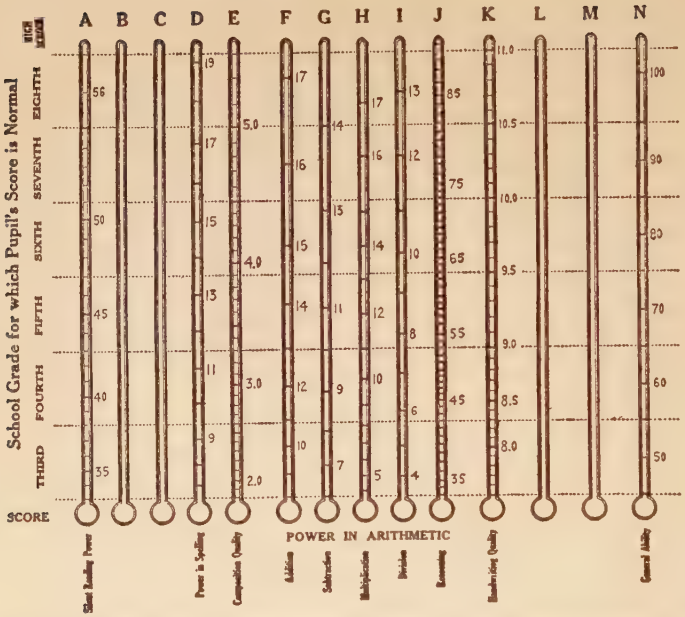


FIG. 1.—Educational Achievement Chart (Eight-Grade-School Edition). Prepared and distributed by M. R. Trabue, Director of the Bureau of Educational Research, University of North Carolina, Chapel Hill, N. C.

DIRECTIONS: Record in the bulb of each thermometer the score obtained on the test which it represents. Then draw a bright colored line (red or blue) upward through the stem of the thermometer to the point corresponding to the size of the score. By following the horizontal dotted lines to the left, one may readily see for what part of the elementary school course the recorded score is normal. Use one chart for the scores of each pupil and one for the average scores of each class or grade. The tests represented above are as follows:

A. T. Scale values on Thorndike-McCall Reading Scales.
B. .....
C. .....
D. 25 word, H-AF list, Buckingham-Ayres Scale.
E. Hillegas English Composition Scale.
F-I. Woody Arithmetic Scales, Series B.
J. Buckingham Scale of Reasoning Problems.
K. Thorndike Handwriting Scale.
L. .....
M. .....
N. Mentimeters.

because the children under different teachers may be unusually bright or dull or may have been exceedingly well or poorly taught, and thus their marks may vary from this general distribution. The distribution of the marks of a large group of pupils will ordinarily follow the normal curve of distribution. Good-sized classes will approximate this distribution. The

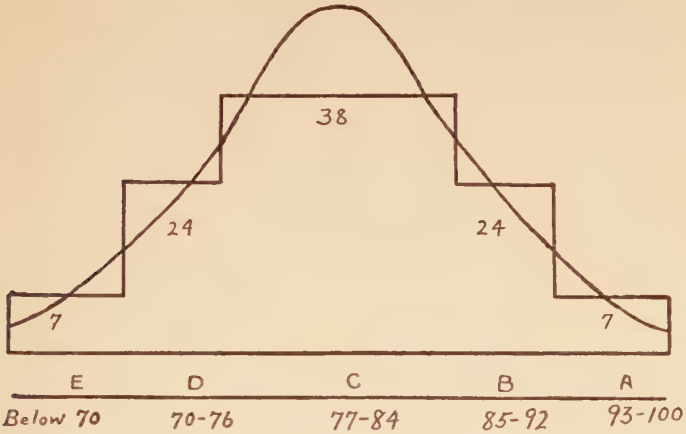


FIG. 2.—The Distribution of School Marks on a Five Point Scale in Comparison with the Normal Distribution Curve.

normal distribution curve is not ordinarily found with very small classes. Teachers can add much to the accuracy of marking if their marks conform approximately to some such standard of distribution as the normal curve. Conformity to such a standard will do much to prevent the "failure set" of some particular teacher. It will also prevent the failures in any particular subject from becoming excessive.

The continuous curved line of Figure 2 represents the normal distribution curve. The rectilinear line represents the normal distribution of marks when five marks are used. The numbers in the diagram represent the number out of 100 that would receive the various marks E, D, C, etc., that are indicated at the

bottom of the chart. The marks are also interpreted in terms of the per cent system of marking, i.e., below 70, 70 to 76, etc., at the bottom of the chart.

An inexperienced principal of a six-six plan consolidated school was recently confronted with the following situation. The pupils came from the sixth grade into the junior-senior

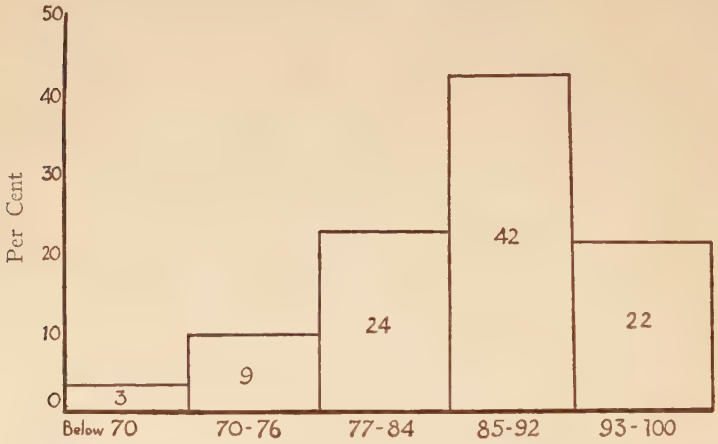


FIG. 3.—Distribution of Marks Given by Fifth and Sixth Grade Teachers.

high school with high marks for their work in the elementary schools. But in the junior-senior high school their first report cards sent home to parents had marks on them much lower than the pupils were accustomed to receive. As a result parents came to the principal with adverse criticism of the junior-senior high school teachers. The principal made a study of the distribution of the marks of fifth and sixth grade teachers and compared them to the marks given by the seventh and eighth grade or junior-senior high school teachers. The distribution of the marks is given in Figure 3 and Figure 4.

This study was then followed by the application of a com-

posite achievement test that revealed slightly lower than standard achievement in all of these grades but a distribution of achievement not very different from a normal one. This lack of coördination in marking between the fifth and sixth grade units of the school and the junior high school unit was largely due to two things: (1) failure to recognize the general rule of distribution of school marks, and (2) a lack of agreement as to standards to be applied in marking.

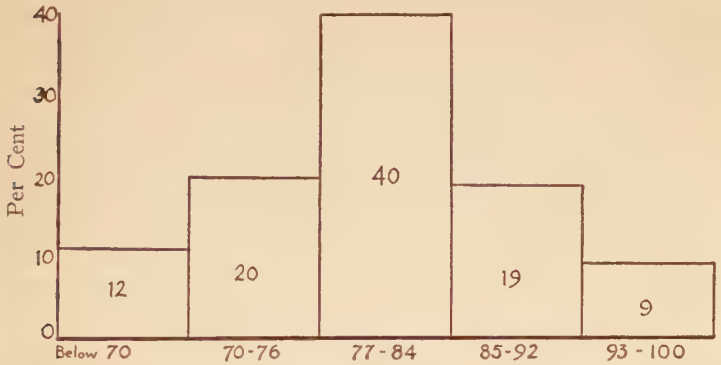


FIG. 4.—Distribution of Marks Given by Seventh and Eighth Grade Teachers.

Teachers' estimates of pupils' ability, or abilities, and traits, when divorced from achievement in school subjects, are valuable. Such estimates by teachers usually show a higher correlation with the results of intelligence or achievement tests than do the actual marks given by teachers. The teachers' estimates of abilities and traits are valuable especially in prognosis and guidance. This value becomes increasingly greater in proportion as the following conditions underlie the estimating: (a) that marks represent only achievement, (b) that the scorer be skilled in marking, (c) that the ratings be based on comparable items, and (d) that these items be as objectively defined as possible. The teachers' estimate of traits and ability

added to other defined factors and measures increases the validity of measures of prognosis to a marked degree.[3]

The traditional written examination or essay-type test has been widely used as a basis for marks and promotion. Its faults have recently been brought before educational workers, and its deficiencies pointed out. These deficiencies are quite generally listed as (1) unreliability, (2) lack of value for diagnostic purposes. The unreliability of essay-type examinations is due to several factors, many of which could be remedied. One of these factors is the limited number of questions, which makes necessary a rather restricted sampling of the whole field of work studied. This weakness is especially undesirable in a final examination. At best only a dozen or fifteen items of importance can be brought into a traditional essay-type examination forty minutes in length. If the test covers a semester's work in the subject, the sampling of questions is of necessity limited. In many cases the sampling is decidedly biased. Two examples are given of examinations that represent on the part of the examiner either a biased viewpoint in sampling or a careless construction of the test as well as a limited sampling.

UNITED STATES HISTORY[4]

(Any six)

1. Contrast the New England Colonies with those of the South, taking Massachusetts and Virginia as types.
2. How was the new constitution different from the old Articles of Confederation?
3. What effect did the Erie Canal have on life in the West?
4. What was the cause of the panic of 1837? Of 1873?
5. Which was more important, the Monroe Doctrine or the Missouri Compromise? Give reasons for your answer.

[3] C. W. Fleming, *A Detailed Analysis of Achievement in the High School.* Teachers College, Columbia University, Contributions to Education, No. 196, 1925.

[4] This test is supposed to cover the whole field of American History usually given in the seventh and eighth grades.

6. What is the "spoils system"? Is it a good policy? Why?
7. What is meant by "Presidential Electors"? How many does Indiana have? How are they chosen?
8. What was the Stamp Act? Why did the Americans object to this act?

GEOGRAPHY [5]

(Any six)

1. Name the States touching the Great Lakes.
2. Name the Gulf States.
3. Trace a vessel from New York to Galveston. What will it possibly take to Galveston? What will the return cargo likely be?
4. Name some ways in which mountains affect climate. Give one illustration.
5. Name and locate four rivers important for commerce.
6. Locate Australia. Name one seaport. Give three products exported from Australia.
7. Name and locate three island possessions of the United States.
8. Name and locate relative to Indiana the states bordering it.

From these lists of questions, with a choice of six out of eight, the pupil could write a perfect paper in Geography without having studied anything but the Geography of the United States and its possessions. In History, by proper choice, the pupil could write a perfect paper without much knowledge of American History past the Civil War. Inadequate and biased sampling not only tends to make such a measurement of achievement unreliable but is likely to influence teachers to emphasize phases of the work inconsistent with the statement of minimum essentials of the subject itself and the objectives toward which the work should be directed.

Ambiguity in the statement of a question, causing lack of

[5] This test was to cover the whole field of Geography usually given in the fifth, sixth, and seventh grades, or the sixth, seventh, and eighth grades.

understanding on the part of the pupil as to what he is expected to do, is an outstanding deficiency in all types of tests. "With whom did the United States fight in the World War?" was a question given on one examination. The meaning of this question was perfectly clear in the mind of the teacher. About 45 per cent of the pupils answered "Germany" and 55 per cent answered "The Allies."

The marking of the papers of the traditional examination is subjective and consequently inaccurate. The marking of a paper by one teacher will not give the same result as would the marking of the same paper by another teacher equally well qualified in the same type or grade of school work. Neither will the same result be obtained if the same teacher should mark two papers of the same pupil, the two papers being based on different sets of questions. When both the teacher and questions are changed, the accuracy of marking is decidedly low. This subject, the marking of papers, has been investigated by Starch, Elliott, Johnson, Wood, Ruch, and others with somewhat uniform results. The following statements adapted from Ruch,[6] who studied the marking of questions in the field of social science, make the situation vivid.

Situation I

Suppose that State Y appointed Miss A to grade the papers of the traditional essay-type of examination for the year 1923. Assume further that Miss A was taken ill and Miss B was appointed as her substitute. If Miss B should grade the papers exactly as Miss A would have graded them, the results would not have been affected by the change in graders. But Miss B could not grade exactly as Miss A would have graded. Consequently the change of graders would affect the results. The amount of influence which this change of graders might have had on the marks of the pupils is measured in a general way by saying that such a method would give results only *20 per*

[6] *Objective Examination Methods in the Social Studies.* Scott, Foresman and Company, pp. 16-18 (1926).

cent better than that arising from the assignment of both sets of marks by pure chance. The fact that Miss B, equally well trained and experienced as Miss A, graded the papers rather than Miss A, reduced the consistency of results 80 per cent. However, had a purely objective examination been used, Miss B's scoring would probably have been exactly the same as Miss A's scoring would have been.

Situation I probably does not occur often in actual practice, but it surely does occasionally. There can be little doubt that errors of relatively enormous magnitude are introduced as a consequence of the "personal equation" of the grader or scorer.

Situation II

Suppose that State Y appointed the same grader for two successive years (e.g., 1923 and 1924). Different sets of questions were prepared for each year. Assume further that a pupil, Franklin Brown, was taken ill during June, 1923, and was forced to postpone his examination until June, 1924. The mark given him would be only *approximately 10 per cent better than assignment of marks by pure chance.*

It has been suggested that Situation II will, in general, yield less reliable results than Situation I, since the former introduces the effects of limited sampling together with those of subjectivity (although the degree of subjectivity is less here).

"The data ..... established beyond serious question that present examinations suffer greatly through the limited sampling possible in 5 and 10 question examinations of the so-called 'intensive' type."

Situation III

This situation is the typical one in actual practice. The several states or other educational bodies which set examinations almost invariably change both the questions and the grader from year to year. If our hypothetical pupil, Franklin Brown, were taken ill in 1923 and forced to wait until the next regular examination period in June, 1924, both the examination and the grader would have been changed.

The general effects upon pupils' marks under Situation III would *be a degree of accuracy roughly 7½ per cent better than chance assignment of marks.*

This unreliability of traditional examinations contributes in a marked degree to the unreliability of teachers' final marks, as the final examination often figures largely in determining the teachers' measure of the pupils' work.

Another defect in the traditional examination is that it causes undue labor for both teacher and pupil. It is doubtful whether for the teacher much time or labor is saved by use of objective or new-type tests if she has to develop the tests carefully herself. The time saved in marking new-type tests over the traditional type is pretty largely offset by the extra time needed for the careful making of the new-type test. For pupils, the new-type test is more satisfying, less grilling, and less laborious than the traditional examination.

An ideal test should indicate the success achieved, and should also serve as an aid to better teaching by pointing out the special needs of the group tested. In diagnosis of pupil needs and difficulties the traditional examination is of very little value. Due to limited and biased sampling and to the composite nature of the answer, it is not an easy task for the examiner to determine the specific difficulty encountered by the pupil. Much attention is now being given to the results of testing as a guide to reteaching a whole class or to meeting the specific difficulties of individuals. The traditional examination offers little assistance in attaining either of these goals.

The value of traditional or essay-type examinations as a measure of the ability of a pupil to organize and express his thoughts is suggested as a consideration for their continued use in a limited degree. In serving this purpose, however, they tend to become more of a teaching device with a value comparable to that of the written composition.

The objective or new-type examination is coming into wide usage, and is discussed in detail in Chapter XXI. The objections made against the traditional examination or essay-type test are overcome by this new type of examination. The new-type test is objective, provides for wide and comprehensive

sampling, and is more satisfying to the pupil. The objections to it are that it does not give any measure of the ability to organize and express one's thoughts, that it may be carelessly made, that it often puts too much emphasis on detailed and unimportant phases of the work, and that the results are not meaningful since they are not comparable with those obtained in other classes and schools. Although these new-type tests will be discussed in a later chapter, it is worth while to say here that they offer an opportunity for more accurate informal testing and more detailed diagnosis than does the traditional examination or essay-type test; accordingly their use should help to make school marks more truly achievement marks.

The standardized achievement test attempts to do away with many of the difficulties noted in the traditional examination and the new-type test.

The method of developing a standard test is fully discussed in Chapter XXII and will only be touched upon here. The content of the test should be and usually is based on the content and expected outcome of instruction as revealed in textbooks and in modern courses of study. The content of a test of this type has been subjected to critical analysis and has had a try-out with the type of pupils who are to use it later. It has met certain statistical requirements of development. It is standardized in respect to giving, timing, and scoring. Its results are in terms of average practice, i.e., norms. It has been validated. Many standardized achievement tests offer much opportunity for diagnosis.

The standard test has many faults inherent in it and is subject to many abuses. (1) The test does not measure everything that it may be supposed to measure. Quite often a so-called language test is a test of technical grammar only. Such a test may not in any sense measure the ability of a pupil to speak. There is a great dearth of tests that measure all the phases of the results of teaching in any one subject. This is of course admitting, as has been said before, that many phases

of achievement in life are not yet measurable. (2) The results of tests are qualified by constant and variable errors.[7]

While standard tests have objective methods of scoring, nevertheless various errors of scoring are possible and probable. Pintner[8] reported this fact in the scoring of intelligence tests, and Monroe[9] in the scoring of reading tests. Various errors are possible and probable in certain types of such tests. This fact is revealed by the experience of almost any one who handles scoring squads or classes. Much of this weakness can be overcome by making more careful and simple directions concerning scoring and by carefully training scorers. It will hardly be entirely removed, however.

Repetition of a test will materially increase the score made by the pupil. This fact has been proved by retesting the same pupils with the same test. It is to be expected then that, should the child be acquainted with the test, his score would not be accurate. The coaching of pupils in the content or special technique of a test will tend to increase the score to the point of making it an inaccurate measure of achievement. A teacher recently told her class that they were going to have a new kind of test on a certain date. The children were very curious and asked her to tell them what it would be like. The teacher read them several examples from each part of the composite test which she gave them later. This reading was helpful to the class as was revealed by the total scores made on the test. Such out-and-out coaching is seldom done, but teachers often unconsciously emphasize certain phases of work to the exclusion of others because they know these phases will be represented in the test. Any results based on tests thus prepared for will tend to be inaccurate.

[7] W. S. Monroe, *The Constant and Variable Errors of Educational Measurements.* Bulletin 15, Bureau of Educational Research, College of Education, University of Illinois.

[8] "The Scoring of Group Intelligence Tests," *Twelfth Conference on Educational Measurements,* Indiana University, 1925.

[9] *Op. cit.,* p. 37.

Intelligence tests are discussed in Chapter XIX. As a measure of ability to learn, they are closely associated with school achievement and school marks. Probably most pupils seldom work up to their limit of ability to achieve. In terms of their ability to do, however, usually those with the least ability accomplish relatively more in our present school situation than do those with greater ability. It is sometimes suggested that marks should be given in terms of what is achieved by those of nearly equal ability. By some people it has been thought desirable to mark the achievement of pupils in accordance with their ability to do. That is, the school should have more than one standard for pupils' work. Let the achievement mark be based on the relation of the child's achievement to his ability to do the work of the given subject. The only evidence of this idea being carried into practice is in schools with special low-ability classes or schools in which the grades are divided into three sections on the basis of ability. In these classes the teachers would still have the normal distribution of marks. An A mark in a poorer section would be designated as A_z, meaning that the pupil made an A mark in a poor section. In the same way an A_x would mean an A in a good section of the grade. The desirability of this practice has not been entirely proved. However, the plan has certain features that are commendable.

BIBLIOGRAPHY

Anderson, David Allen. "Methods of Measuring Teaching Efficiency," *School and Society,* Vol. 3: pp. 556-562, April 15, 1916.

Blackhurts, J. Herbert. "The Normal Curve As Related to High School and College Grading," *School and Society,* Vol. 13: pp. 447-450, April 9, 1921.

Boyce, A. C. "Methods for Measuring Teachers' Efficiency," *14th Yearbook National Society for the Study of Education,* Part 2, 1915.

Camp, F. S. "Some Marks: An Administrative Problem," *School Review,* Vol. 25: pp. 697-713, December, 1917.

Cattell, J. McK. "Examinations, Grades, and Credits," *Popular Science Monthly,* Vol. 66: pp. 367-78, February, 1905.

Dickinson, Z. C. "Suggestions Toward Improving Examination Marks," *Bulletin University of Minnesota,* Vol. 26: pp. 31-36, August, 1923.

Eaton, H. T. "A Study of School Marks," *Education,* Vol. 43: pp. 620-628, June, 1923.

Harrer, E. "Reflections on a New Method of Grading," *School and Society,* Vol. 24: pp. 582-584, November 6, 1926.

Hillbrand, E. K. "A High School Marking System," *School and Society,* Vol. 21: pp. 142-43, January 31, 1925.

Honeywell, R. J. "A Convenient Grading System," *Educational Review,* Vol. 68: pp. 26-29, June, 1924.

Horrall, A. H. "Uniformity in Grading," *School and Society,* Vol. 22: pp. 594-95, November 7, 1925.

Klapper, Paul. "Standards of Measuring and the Differentiated Drill," *School and Society,* Vol. 4: pp. 671-72, October 28, 1916.

Korey, Ruth A. "The Question of Marks," *School and Society,* Vol. 24: pp. 175-76, August 7, 1926.

Lockhart, A. V. "High School Marking Systems," *School and Society,* Vol. 21: pp. 409-10, April 4, 1925.

Mersereau, E. B. "An Experiment in Marking," *School Review,* Vol. 34; pp. 348-56, May, 1926.

Odell, C. W. "High School Marking Systems," *School Review,* Vol. 33: pp. 346-54, May, 1925.

Ohlson, David. "School Marks Versus Intelligence Rating," *Educational Administration and Supervision,* pp. 90-102, February, 1927.

Pressey, S. L. "Fundamental Misconceptions Involved in Current Marking Systems," *School and Society,* Vol. 21: pp. 736-38, June 20, 1925.

Roberts, Alexander C. "Measuring and Testing in Education," *The Journal of the National Education Association,* Vol. 13: p. 101, March, 1924.

Rogers, James H. "A Uniform Grading System," *School and Society,* Vol. 22: pp. 160-63, August 8, 1925.

Rugg, H. O. "'What is Credit for Quality?' A Summary and Evaluation of Published Discussions," *Elementary School Journal,* Vol. 19: pp. 634-37, April, 1919.

Santee, J. F. "Mental and Educational Tests in Relation to Teachers' Marks," *Education,* Vol. 47: pp. 155-63, November, 1926.

Shearer, W. J. "The Elizabeth Plan of Grading," *Addresses and Proceedings of National Education Association,* Vol. 37: pp. 441-48, 1898.

Somers, Grover T. "A Proposed Marking System for Colleges and Universities," *Bulletin of the School of Education,* Vol. 3, No. 6, Bureau of Coöperative Research, Indiana University, School of Education.

Somers, Grover T. *Students' Attitude Toward Examinations,* Bulletin of the School of Education, Vol. III, No. 1, Bureau of Coöperative Research, Indiana University.

Tidyman, W. F. "Adjusting Marking Systems to Differences in Groups," *School and Society,* Vol. 22: pp. 247-48, August 22, 1925.

Waller, O. L. "How Shall We Measure Scholastic Attainment," *School and Society,* Vol. 22: pp. 818-20, December 26, 1925.

CHAPTER III

Criteria for Judging Standardized Tests

Teachers, supervisors, and administrators are constantly facing the question of what tests to use. Many universities and Bureaus of Research have hundreds of letters each year asking advice on this question. It is vital that those having the responsibility of interpreting test results should have adequate guidance in selecting the tests to be used. The following letter is an example of the questions often raised by those who wish to introduce a testing program.

August 27, 19...

Bureau of Coöperative Research,
School of Education,
Indiana University,
Bloomington, Indiana.

Dear Sir:

I am planning to classify the pupils of grades one to six of my school for instruction purposes. At the same time I want to select, by testing, about a score of deficient or subnormal pupils for a special room. What test would be required for the purpose of classifying the pupils in the various grades? What test should be used for selecting deficient or subnormal pupils? There will be in all about 60 to 80 pupils in each grade and a total of not more than 400 children in the school.

I also wish to do some testing with the deficient group to find out special difficulties. What diagnostic tests do you recommend in Arithmetic and in Reading?

I also wish to give some achievement tests during the year to the group of normal pupils to determine if their work is up to standard. Will you recommend some tests for this purpose?

Will you give me an estimate of the expense involved in such a testing program as I have roughly outlined? It is not easy

for me to get money for this kind of work. Therefore, I shall need a rather careful estimate beforehand on the total cost of the tests as well as some estimate of the time that it will take to do the work connected with their use.

Very truly yours,

Principal Consolidated Schools.

P.S. If you do not sell these tests, please send me the address of the publishers.

The three-page answer to this letter was gladly written, since one of the purposes of the bureau is service to the school people of the state.

It was doubtless in response to a general demand for guidance in test selection that Arthur S. Otis recently published a scale for rating educational tests.[1] This scale is reproduced on page 34.

The scale contains the major points that have to be considered in the selection of a test. Certain vital points should be raised in connection with it.

I. Is the test valid?

A test is valid if it measures what it purports to measure. If it is an authentic test and is supposed to measure the legitimately expected outcomes of the teaching of a certain subject, then it is fair to ask, "Is it consistent with the objectives of education, with the objectives of the subject, with the content of the course of study, with the best practices in teaching?"

It is further necessary to ask, "Does it measure adequately the knowledge, traits, skills, functions, attitudes, or content of the school subject that it is supposed to measure?" A test in arithmetic may not propose to measure all the outcomes of instruction in that subject. If it proposes to measure one outcome, does it measure that outcome adequately? If the test

[1] *Scale for Rating Tests,* Test Service Bulletin No. 13, World Book Company, 1926.

Scale for Rating Tests	*Names of Tests*				
Manual (5)					
Validity (15)					
Reliability (10)					
Reputation (5)					
Ease of Administration (Total 15) (a) Preparation (4)					
(b) Time limits (4)					
(c) Explanation needed (3)					
(d) Alternative forms (4)					
Ease of Scoring (Total 15) (a) Objectivity (10)					
(b) Time required (3)					
(c) Simplicity (2)					
Ease of Interpretation (Total 15) (a) Norms (5)					
(b) Directions for interpreting (4)					
(c) Class record (1)					
(d) Application of results (5)					
Convenient Packages (5)					
Typography and Makeup (5)					
Test Service (10)					
Total (100)					

proposes to serve a particular function, does it do so adequately? Often tests called diagnostic fail to perform the function of diagnosis to any marked degree. Such a failure makes the test invalid in function because it does not measure in detail the trait or skill to such a point that remedial work can be direct and purposeful.

It would be valuable if each publisher of achievement tests would state which objectives of a subject are tested and which ones are omitted. In the list of objectives of the teaching of spelling this is often given: "To develop the meaning and use of the words to be spelled." Suppose a test of words is taken from the Iowa Spelling Scale and pronounced to children to be written by them. A pupil taking this test makes a score of 84 per cent. This does not accurately indicate either his knowledge of the meaning and use of the words or the degree of success the teacher has had in the teaching of this important phase of spelling. Such a test may be a valid measure of "the ability to spell correctly words most commonly used," but yet not be valid as a test of the meaning and use of the words spelled. However, certain traits and achievements may be measured indirectly by the measure of closely associated or highly correlated traits or achievements.

Achievement tests should be validated as to content by a check that would insure their paralleling good courses of study and much used textbooks and including results of scientific studies in education, especially in the field of curriculum. The judgments of experts and practitioners are invaluable in the development of this phase of a test.

Two samples of this type of validation of standard tests will show what is meant.

(1) Course-of-study type and textbook type.

GATES-STRANG HEALTH KNOWLEDGE TEST

"The Health Knowledge Test is based upon a careful analysis of the facts concerning many phases of healthful living which

were found in twenty selected courses in health offered in rural and city schools in America, and in fourteen of the most widely used textbooks on health."

(2) Examination and expert opinion.

RUCH-COSSMAN BIOLOGY TEST—MANUAL OF DIRECTIONS

1. "A letter was sent to the Superintendent of Public Instruction in every state of the Union having a central officer, asking that he list the names of those schools and teachers who were doing the highest grade work in high school biology. Lists were received from 23 states, with a total of 126 schools answering to our specifications.

2. "Each of the 126 teachers was asked to send duplicate copies of all final examination questions used during the current year. Lists for previous years also were requested, and in some cases these were available for periods of five years or more prior to 1918. The total number of questions received ran slightly over two thousand.

3. "The two thousand questions were then analyzed for frequency and condensed and classified as 300 constantly recurring items.

4. "The final list of 300 questions was then submitted to 100 of the leading teachers and authorities on the teaching of biology, for ratings as follows. . . . Provision was also made for the writing in of any other questions not already listed. . . ."

II. Is the test reliable?

In measuring with a yard stick, one may expect for all practical purposes to get the same results at one time as at another. Such uniformity of results should be expected also from the use of scientific instruments of measurement in education. Reliability is the term used to designate such a quality in standard tests, and represents the accuracy with which the test measures whatever it does measure. Used today or tomorrow, the test with perfect reliability will give the same results; or, if used both today and tomorrow, it will give the same results except for a difference resulting from practice ef-

fect. The measure of reliability is usually designated by the coefficient of correlation secured between the scores made by the same pupils taking the test twice the same day or between the scores made by the same pupils on equivalent forms of the same test. This measure is also obtained by correlating the total scores of the same pupils made on two applications of the same test. A second method of obtaining this measure is correlating the scores of the same pupils made on even items of a single test with the scores made on the odd items of the same test. It is necessary to carry out certain statistical procedures of correction if the second method is used. This method is more fully explained in Chapter XXII.

The reliability of a test is important and should be stated in terms of the group with which the reliability was estimated, for the reliability of a test is dependent upon the group to which it is given. For example, the Thorndike-McCall Reading Scale has a correlation of .75 using pupils of Grades IV-VIII and of .60 using only Grade VI. The significance of a reliability coefficient is somewhat debatable; however, the following statement may be generally accepted.

A reliability coefficient ranging from

.95 to .99 is seldom attained in standard tests

.90 to .94 is highly desirable in testing

.80 to .89 represents some of the best tests now on the market

.70 to .79 represents many tests now on the market

.69 and below is not satisfactory for individual measurement. Tests with such a reliability coefficient are used in group testing.

There are various factors which affect the reliability of a test. Some of these can be observed by any one judging a test. The length of a test has much to do with its reliability. The longer a test, other things being equal, the more reliable it will be. The total score of an omnibus or composite test is likely to be reliable because it represents more than two or three hours

of work on the part of the pupil. It is necessary, then, to look very carefully at the reliability of a test requiring only five minutes of the pupil's time.

A source of unreliability, which is often outside the control of the test maker, is found in the conditions under which the test is given. Much excitement on the part of some pupils, due to the game they have been playing in the gymnasium or to a noise of a disturbing kind outside the building, may change the results of the test.[2] This probably explains why the reliability often found when the test is used in laboratory schools or scientific laboratories is not always attained in field work. Pupils change from hour to hour in their reactions.

Instructions for giving and scoring a test have very much to do with the reliability or consistency of results obtained from it. Instructions that are detailed and specific and that allow for only one method of procedure to be followed by everybody, are the ones which make for the highest reliability. The examiner may be a source of unreliability in results if he fails to carry out the instructions and conditions of the giving and scoring of the test.

The reliability of tests is discussed in detail in Chapter XXII.

III. Administration of Tests

It is highly important that a test have careful instructions both for the examiner and for the pupils. The aim of instructions in administration of tests is primarily to get all pupils to understand what is to be done and to have them perform in a uniform manner under standard conditions. Everything that contributes to this end is of value to the scientific measurement of achievement. Instructions to pupils should be printed both on the blanks for pupils and on those for examiners. Instruc-

[2] Symonds holds that the influence of these factors is exaggerated and that there are seldom differences in scores on a test as a result of such factors. P. M. Symonds, "A Study of Extreme Cases of Unreliability." *Journal of Educational Psychology*, Vol. 15, pp. 99-106, Feb., 1924.

tions to the examiner should appear only on the examiner's sheet.

EXAMPLE OF INSTRUCTIONS FOR GIVING THE TEST

The following is an example of good instructions for the giving of a test.

Before distributing the test booklets, have the desks cleared and see that each pupil has a sharpened pencil. A number of extra sharpened pencils should be available so as to supply pupils who break their pencil points. Begin the test by saying, "We shall now pass out some test booklets. Do not write on the books nor open them until I tell you to." If the pupils are in the junior and senior high school and have previously taken standard tests, this instruction can be changed to "as soon as you receive the booklets, fill out the blanks on the first page."

Have the tests distributed, one to each pupil, face up, then say, "At the top of the page on the first line you will see the word 'Name.' On that line write your name, your first name first, then your middle name, and then your last name. On the next line it asks for 'Age,' last birthday, in years. Write in that blank how old you were on your last birthday. On the same line you find 'Birthday.' After that write the month and day of your birthday. On the next line it asks for grade or class. Write 5A in that blank. On the same line is the word 'Teacher'; write the name of your teacher in the blank after 'Teacher.' On the next line you find 'Date of Examination'; write the Month, Day, Year. On the same line you will find 'City'; after that write [name of town]."

See that the blanks are properly filled out by moving around among the group. After this is done, go back to the front of the room and there say, "Look at the directions at the bottom of this page—I will read them. This is a test to see how well you have studied your arithmetic. The test consists of two parts. You are to do the first part now. The first part has 40 computation problems. When I say, *'Ready, begin,'* and not before that, turn over the page and begin with the first problem and work them in order. Do as many as you can, but be careful not to make mistakes. Some of the problems are on the third

page. Do not forget to turn the page to them. You may figure on this paper. You will be graded on the number of problems you do correctly. [Pause.] Now look at the sample problem—see how it is worked. [Pause.] You will be given 40 minutes for this part of the test. When I say *'Stop,'* you should immediately stop working and close your booklets. If you should finish before time is called, look over your problems and be sure that they are all correct. Ready, begin!" At the end of exactly 40 minutes say, "Stop. Close your books and rest a few minutes."

Other items not in the example above are often found to be valuable in test instructions. Examples of such suggestions are: (1) "Do not spend too much time on one question; if you can't do it, leave it and go on to the next; come back to it later." (2) "When you have finished Part I, go right on to Part II."

SCORING THE TEST

If a test is to be scored accurately and in the same way by different people, there must be no doubt concerning the correctness or incorrectness of the answers indicated or given by the child. A test is objective when it can be scored in this manner. The more objective the test, the more likely it is that the same test will be scored the same by two different people.

The scoring of a test should be made as easy and simple as is consistent with good results. The scoring key and directions for scoring should be specific and definite. Problems are constantly arising concerning part answers, erasures, alterations, marking several items in one answer, and omissions. These should be taken care of for the scorer in the instructions concerning scoring and in the scoring key. Specific instructions, with examples where necessary, on how to obtain the score should be included. The faster and more easily a teacher can handle the scoring of tests, the more energy and time she will have for the intelligent interpretation and diagnosis of results

and for making plans for remedial instruction. Cumbersome methods of scoring should always be avoided unless they add very materially to the value of the test. "How many tests can be scored per hour?" is therefore one of the questions that should be asked concerning any test to be bought.

INTERPRETATION OF RESULTS

The results obtained from the scoring must be interpreted by comparison with some standard. This standard, called a norm, is an average to which measures are compared. A standard test implies norms. These norms should be based not only on a large number of cases or pupils but upon a representative group well sampled from the types of situations like those in which the test is expected to be used. The norms should be expressed not only in terms of the number of pupils but also in terms of the kind of group they represent. An example of the type of statement that should accompany any table showing norms is as follows:

Table of Norms on Indiana Composite Achievement Test Form *A* for Grades Seven and Eight. These norms are based on the score of 15,280 cases from one-room, two-room, three-room, and consolidated rural schools, May, 1927.[3]

For the purpose of interpreting scores, various types of norms are given with tests, such as Grade norms, Age norms, Percentile norms, and T-Score norms. The more methods of interpretation of results that are given, the more meaningful the scores can be made if proper directions for their interpretation accompany the test. This subject is further discussed in Chapter XXII.

The Manual of Directions is a valuable part of any test and deserves higher rank than is given it by the rating scale shown

[3] Forms A, B, C, and D of the Indiana Composite Achievement Test are published by Bureau of Research, School of Education, Indiana University.

on page 34. The Manual should include such items as (1) Name of test and authors, (2) Purpose of test, (3) Basis for the construction of the test, (4) Description of the test, (5) Standardization of the test, (6) Validity of the test, (7) Reliability of the test, (8) Directions for giving the test, (9) Directions for scoring the test, (10) How to interpret the results in terms of as many type norms as would be usable, (11) Uses of the results in various phases of school work, such as classification, diagnosis, administration, supervision.

While the cost and mechanical makeup of a test are important and have much to do with its use, they are not to be ranked in importance with many other factors mentioned in this chapter as criteria for selection of tests.

BIBLIOGRAPHY

Campbell, A. L. "Keeping the Score," *School Review,* Vol. 29: pp. 510-19, September, 1921.

Franzen, R. and Knight, F. B. "Criteria to Employ in Choice of Tests," *Journal of Educational Psychology,* Vol. 12: pp. 408-12, October, 1921.

McCall, W. A. "Criteria for Determining Which Educational Test to Employ," *School and Society,* Vol. 7: pp. 207-09, February 16, 1918.

Russell, Charles. *Classroom Tests,* Part II, Chapters IX, XIII, XIV, XVII, Ginn and Company, 1926.

Symonds, Percival M. "A Study of Extreme Cases of Unreliability," *Journal of Educational Psychology,* Vol. 15: pp. 99-106, February, 1924.

Wilson, G. M. "Criteria of a Standard Test," *Educational Review,* pp. 138-41, March, 1926.

CHAPTER IV

Interpreting and Reporting Scores

Giving the Test

If the manual of directions accompanying a test meets the requirements set forth in the preceding chapter and if the instructions in the manual are rigidly adhered to in giving the test, there is not much need for additional specific directions on any one test. The following additional suggestions of a general nature, however, will be helpful in the administration of most tests.[1]

(1) The person giving the test should study the directions carefully. He should practice giving the test to some person or persons if possible, but at least he should read the directions aloud several times. The need for familiarity with the directions can not be overemphasized.

(2) The room should be quiet throughout the test. Strict obedience and attention should be demanded. No questions should be allowed during the test. The person giving the test should speak distinctly and at a moderate speed, and cultivate a manner which is agreeable but which at the same time suggests authority. Pupils should be made to feel "at home" in taking the test. They should look forward to taking the test without fear or nervousness if the test is properly given and if no misconceptions with regard to the meaning and use of the results are allowed to arise.

Throughout the examination, commands should be given in a clear and energetic voice. The directions should be given

[1] Harry A. Greene and A. M. Jorgenson, *Using and Interpreting Tests.* University of Iowa Extension Bulletin Number 139, College of Education; Series No. 15, January, 1926, pp. 17-18. (The material is adapted.)

slowly and with careful attention to emphasis. The voice should be just loud enough to carry to all parts of the room. The directions for the tests should be followed verbatim.

(3) In timing the test, a stop watch is desirable. If an ordinary watch is used, it is necessary to exercise great care to secure accurate timing. In such cases it is well to have an assistant whose sole duty is to act as timekeeper for the test. If this is impossible, the following method will serve reasonably well:

	Hr.	*Min.*	*Sec.*
(a) Record time (on a paper or on the blackboard) starting signal is given	11	18	20
(b) Add to this the time required for test		5	00
(c) Total—time that stop signal is due..	11	23	20

(4) The scoring of most standard tests is made almost wholly objective by the use of scoring keys. The keys and directions for scoring each specific test should be followed scrupulously, even though they may seem contrary to one's own judgment.

Interpreting the Scores

After a test has been given and scored, the teacher must interpret the scores. The first thing to do in preparation for this interpretation is to tabulate the scores. Tabulating is performed as follows:

Suppose you have given a test to the sixth grade pupils in arithmetic in your room, and the scores are as follows:

A.B. 15	A.L. 15	A.W. 17	B.G. 16
A.C. 16	A.M. 10	A.X. 15	B.H. 18
A.D. 14	A.N. 21	A.Y. 15	B.I. 15
A.E. 15	A.O. 13	A.Z. 15	B.J. 17
A.F. 15	A.P. 22	B.A. 14	B.K. 14
A.G. 16	A.Q. 16	B.B. 17	B.L. 14
A.H. 13	A.R. 19	B.C. 13	B.M. 14
A.I. 10	A.S. 14	B.D. 16	B.N. 9
A.J. 15	A.T. 18	B.E. 15	B.O. 15
A.K. 14	A.U. 16	B.F. 14	

Look through the papers and find the one with the highest score; then find the paper with the lowest score. The difference between the two gives you the *range* of scores, which in this case is from 9 to 22. Then, as is indicated in Table 1, tally

TABLE 1.—DISTRIBUTION OF PUPILS' SCORES MADE IN WOODY-McCALL ARITHMETIC TEST. FORM I, 5B CLASS, SCHOOL X, CITY Y, 1927

Score	*No. Pupils Making (frequency)*	*Total (frequency)*
22	|	1
21	|	1
20		0
19	|	1
18	||	2
17	|||	3
16	𝍸 |	6
15	𝍸 𝍸 |	11
14	𝍸 |||	8
13	|||	3
12		0
11		0
10	||	2
9	|	1

the scores in the usual way under column headed "No. Pupils Making (frequency)." In the next column, headed "Total (frequency)," write the figure that represents the total number of persons making each score. It is well to write in the zeros to prevent confusion. The scores are now tabulated and offer a sort of picture of the situation in the class so far as the tests reveal the situation. The two pupils that made 22 and 21 are

far above most of the group. The two pupils that made a score of 10 and the one that made a score of 9 are far below most of the group.

This example involves only a small group of children, and the scores are rather small. Let us take an example with a larger range of scores and with more cases. The following are total scores on the Indiana Composite Achievement Test where all the children of the eighth grade of the county were tested. The following scores were made by children in X County:

824, 778, 775, 758, 375, 278, 426, 452, 406, 418, 421, 724, 431, 751, 748, 728, 678, 690, 684, 691, 716, 724, 716, 700, 711, 456, 464, 474, 543, 526, 531, 529, 462, 471, 528, 494, 678, 654, 659, 673, 669, 698, 370, 578, 681, 650, 666, 671, 664, 500, 550, 523, 521, 516, 509, 518, 511, 649, 625, 640, 631, 629, 636, 600, 628, 609, 601, 621, 616, 612, 528, 539, 550, 560, 570, 555, 576, 598, 588, 587, 584, 579, 558, 564, 571, 566, 567, 571, 558, 562, 564, 566, 300, 858, 591, 498, 581.

The lowest score here is 278, and the highest score is 858. To write down each score between these would be too much of a task, and the sheet on which they were written would be too awkward to handle. It would be well, therefore, to decide on some method of grouping the scores. In determining the grouping to be used, it is usually best to have not more than thirty groups nor fewer than ten. The difference between the highest score and the lowest score divided by 25 or 30 will give a reasonable idea of the size of the grouping to be used. In the scores just given 858 minus 278 equals 580, which divided by 25 equals 23⅕. This result suggests that the unit nearest to this that would be easy to use is 25. In this case 25 is called the class or step interval. The step interval refers to the size of the grouping. For example, in the 575-599 group the class interval is 25; or, again, all of the scores between 525 up to and including 549 are in one step interval. With 25, then, as the unit of grouping, or the step interval, the distribution of the scores would be as found in Table 2, which follows.

TABLE 2.—DISTRIBUTION OF TOTAL SCORES ON INDIANA COMPOSITE ACHIEVEMENT TEST. FORM A, 8A CLASS, COUNTY X, 1927

Scores	*No. Pupils Making (frequency)*	*Total (frequency)*
850–874	|	1
825–849		0
800–824	|	1
775–799	||	2
750–774	||	2
725–749	||	2
700–724	𝍸 |	6
675–699	𝍸 ||	7
650–674	𝍸 |||	8
625–649	𝍸 ||	7
600–624	𝍸 |	6
575–599	𝍸 ||||	9
550–574	𝍸 𝍸 𝍸	15
525–549	𝍸 ||	7
500–524	𝍸 ||	7
475–499	||	2
450–474	𝍸 |	6
425–449	||	2
400–424	|||	3
375–399	|	1
350–374	|	1
325–349		0
300–324	|	1
275–299	|	1

The scores made on the test and listed on page 47 can be tallied in as before. It is somewhat easier, however, to put the papers into 24 piles according to the grouped scores. The number of papers in each pile can then be counted and entered in the distribution table under the column "Total (frequency)" opposite the proper group. Sometimes it is desirable to distribute the test scores made by the pupils of several grades on the same chart. By using the same procedure of tallying scores or counting papers, a result like the one found in Table 3 will be obtained. This gives a crude method of comparing the results of different classes and groups.

The median score of each group is found in the step indicated by the groups of cases that are found in the brackets. For example, in grade 3 the median score is somewhere in the step 25-29; in grade 4 in the step 40-44, etc. By noting this, a rough check on the difference between the scores of the various grades can be made.

The Results Expressed in Some Single Term

Often it is desirable to express the work of a class in some single term. Two terms very often used for this purpose are the *mean* (generally called the average), and the *median*.

To obtain the mean with a small number of scores the easiest way is to add them all together and divide by the total number of pupils. Turn back to the first example of the scores made by the class in arithmetic, page 44. By adding all these scores together and dividing by 39, which is the number of pupils, the mean score is found to be 590 divided by 39 or 15.1. This method of finding a single term as a measure of the central tendency of the scores has two disadvantages; first, it is long and difficult to calculate; second, if a few individuals make a very low or a very high score, the score of the whole class is unduly affected thereby.

To obtain the median, another measure of central tendency, we use another method, in some ways as accurate if not more

TABLE 3.—DISTRIBUTION OF SCORES ON ABX TEST. GRADES 3, 4, 5, AND 6, SCHOOL C, APRIL, 1927

Score	*Grade*			
	3	4	5	6
105–109				1
100–104				2
95– 99				3
90– 94				3
85– 89		3	3	7
80– 84				3
75– 79		1	5	7
70– 74			7	(13)
65– 69	1	3	9	10
60– 64			7	7
55– 59	1	6	(13)	4
50– 54	2	7	5	2
45– 49	2	13	9	3
40– 44	7	(10)	2	
35– 39	8	12	2	1
30– 34	8	8		
25– 29	(12)	2	2	
20– 24	10	4		
15– 19	13	2		1
10– 14	4	4	1	
5– 9	5			
0– 4	4	1		
Total.....	77	76	65	67

so than the method used in finding the mean. Put down all the numbers in the order of value from the highest to the lowest. In the arithmetic test this arrangement would be: 22, 21, 19, 18, 18, 17, 17, 17, 16, 16, 16, 16, 16, 16, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 14, 14, 14, 14, 14, 14, 14, 14, 13, 13, 13, 10, 10, 9. Or, still better, arrange the papers in the order of the values of the scores on them; then count down or up to the middle paper. The score found on the middle paper in this case is 15. The 15 underscored twice is this score which is called the mid-score or median score. This is a type of average which is often used to represent the achievement of a class as a whole. The middle or median score is much more easily found than the mean, and is less affected by extremely high or low individual scores. Wherever the frequency distribution of cases is just like or nearly like the normal curve, these two measures, the mean and the median scores, are practically the same.

Another single term that represents a type of average is the *mode*. The mode is the score made by the most pupils. In the example given the mode would also be 15 because 15 is the score made by the largest number of pupils. This measure is seldom used because it is unstable and may be markedly changed by the shifting of a few cases.

Sometimes it is desirable to find the median score when the material is grouped in a distribution like that found on page 47. The total number of cases here is 97. If the papers were at hand, they could be arranged in order of size of scores and the middle paper found. This would be the median score. If the median score of this group is desired after the material is tabulated in the form of Table 2 and the papers are not available, the median point can be calculated.

To find the median point or mid-point on the range of scores from 278-858 in the distribution shown in Table 2, the following procedure can be used:

Divide the total number of scores, or the frequencies, by 2. The total number of scores or frequencies in this case is 97. Ninety-seven divided by 2 equals 48½. To get this 48½, count or add the frequencies in the frequency column beginning at the bottom, but do not count or add beyond 48½ frequencies. By counting up to and including frequencies in the step interval 550-574, the total is 46. It can be seen that the center point is above 574. If the frequencies of the next group are added in, the total will be 54. It can be seen from this that the median point is not 600 or above. Fifty-four is too many and 46 is not enough, as the number desired is 48½. The middle point in the whole range of scores is, then, somewhere in the group of scores 575-599. To get the exact location of the center point, proceed as follows. Since 48½ is the middle of all the scores and 46 is the number of frequencies or cases up to the group 575-599, then to get the remaining 2½ cases it is necessary to go into that step, 575-599, and take 2½ of the 9 frequencies or cases or scores found there. Now it is assumed that the total 9 cases are equally distributed over the group 575-599. To get the center or median point, then, in this group of scores, take 2½/9 of 25 (the range of the step 575-599). This gives 7. Add this to the 575, which is the lowest score in this step interval; this gives 582. Five hundred and eighty-two is the middle point of the scale, below which 50 per cent of the cases fall and above which 50 per cent of the cases fall. To make this clearer, we may outline each step as follows:

Problem—To find the median point on the scale when the data are grouped in step intervals of 25 (Table 2).

a. Divide total number of cases, or frequencies, by 2: $97 \div 2 = 48\frac{1}{2}$.

b. Add all the frequencies of each step, but do not go beyond one-half the sum of the frequencies, which is 48½. Add (from the bottom) all the frequencies until 46 is obtained; to add another step would be to go beyond the 48½.

c. Call the beginning of the next step the estimated median. 575 is the estimated median.
d. Find the difference between the number of frequencies counted and one-half the total sum of frequencies: $48\frac{1}{2} - 46 = 2\frac{1}{2}$.
e. Find the part of the next step needed to obtain one-half the sum of the frequencies. This step contains 9 frequencies. $2\frac{1}{2}/9 = \frac{5}{18}$.
f. Multiply this by the size of the step interval: $\frac{5}{18} \times 25 = 6\frac{17}{18}$ or 7.
g. Add the result of step *f* to the estimated median: $575 + 7 = 582$. This is the true median point.

Quite often it is desirable to represent the standing of the class not only by a single term like the *median score* or the *mean* but also by such terms as will show how the scores of the class were distributed. One way of doing this is by showing the whole distribution as is done in Tables 2 and 3. A better way is to show the same results in diagram form. It would be possible to show this partly by giving the *range* as in the case of scores on the Indiana Composite Achievement Test, Form *A*, from 278 to 858, or on the arithmetic test from 9 to 22; however, on the whole the range has little meaning.

Another way of indicating the score of a class is by giving the range, not of the whole distribution, but of the middle 50 per cent of the distribution. The lower limit of this middle 50 per cent is called the Q_1, or the *first quartile* score, of the distribution, and the upper limit of the middle 50 per cent is called the Q_3, or the *third quartile* score of the distribution. To find the quartile scores, count up from the bottom one-fourth of the papers or scores or frequencies. Refer again to the case of the arithmetic scores, page 50. Count to the tenth score, which is approximately one-fourth of the papers. The score is 14 (notice that it is marked $\underline{14}$); this is called the first quartile score. The third quartile score is obtained by counting three-fourths of the way through the papers or scores from the bottom up or one-fourth of the way through the papers or

scores from the top down. This score would be 16 (notice that it is marked $\underline{16}$); the middle half of the distribution, then, is from 14 to $\underline{16}$ or from Q_1, equals 14, to Q_3, equals 16.

Still another manner of expressing the distribution of scores is in terms of the *percentile*. This method is most valuable with larger groups. The procedure is as follows: The 25 percentile is represented by Q_1; that is, approximately 25 per cent are below and 75 per cent are above this score. The median, which might be called Q_2, is the 50 percentile; in this case 50 per cent are above and 50 per cent below the score; and Q_3 is the 75 percentile, with 25 per cent above and 75 per cent below it. In the same manner the 10 percentile or 90 percentile can be approximately determined. The more scientific method of determining the percentile is given in Chapter XXII. The percentile score is of much value in the statement of subject norms for junior and senior high school subjects. Percentiles will be more widely used as they are better understood.

The *mean deviation* is sometimes used to indicate how widely the various scores differ from the average or mean score. To find the mean deviation, (1) find the mean, (2) find out how much each score differs from the mean, (3) add all these differences together and divide by the total number of cases. The result is the mean deviation. This represents the average of the differences between the individual scores and the mean score for the group as a whole.

In the preceding chapter it was stated that a norm is used with a test and that it represents an average achievement of some type. It was also noted that such norms as age, grade, percentile, and T-score norms are valuable. An *age norm* is the average of the scores made by children of a certain chronological age. A ten-year age norm represents the average of the scores made by the ten-year-olds taking the test. If such an age norm is given, it should represent the average of all ten-year-old children regardless of the grade in which they are situated. It is for this reason that age norms obtained above

the age of compulsory attendance at school are not likely to be reliable. Many children will have dropped out of school by this time, and thus it would be difficult to measure them. An age norm figured for such a group would not be truly representative. Age norms for the children of the lower grades of the elementary school are, however, valuable.

The age norm is given also in terms of subject age, educational age, and mental age. A *subject age,* let us say ten years, represents the average achievement in a certain subject of representative ten-year-olds. An *educational age* of ten years represents the average achievement on a composite test of representative ten-year-olds. If a pupil makes on mental tests the average score made by eight-year-old pupils on the same tests, he is said to have a *mental age* of eight, no matter what his chronological age may be. In certain respects age norms represent objectively determined differences that grade norms do not show. Many factors of growth are affected mostly by age, and the norms of the age type give the best basis for comparison and differentiation. Tests of physical capacity are better described by age norms than by grade norms.

The *grade norm* is the average of achievement of a certain grade. It is liable to the same misrepresentations as age norms. To be a true norm, it must truly represent all types of a certain grade. These norms are especially valuable with those measures that show the results of teaching or drill. Grade norms need corrections for various times of the year because achievement changes during the school year. A fifth grade norm, for example, would be more useful if stated as a beginning of 5A norm, an ending of 5B norm, or a mid 5B norm.

In connection with the Stanford Achievement Test an even more elaborate system of norms has been worked out for both age and grade. A section of the norm chart of that test is given in Table 4. This norm chart is for the primary examination, Form *A*.

TABLE 4.—EDUCATIONAL PROFILE CHART OF STANFORD ACHIEVEMENT TEST[2] (PRIMARY EXAMINATION)

Test 1, Parag. Mean.	*Test 2, Sent. Mean.*	*Test 3, Word Mean.*	*Read. Total*	*Test 4, Arith. Comp.*	*Test 5, Arith. Reas.*	*Arith. Total*	*Test 6, Dictation*	*Total Score*	*Educational Age*	*Chronological Age*	*Grade*
-49	-28	-31	-108	-76	-42	-118	-74	-30	-10-8	-10-8	-4.7
-46	-27	-30	-103	-74	-41	-115	-72	-29	-10-5	-10-5	-4.5
-44	-26	-28	- 98	-72	-40	-112	-70	-28	-10-3	-10-3	-4.3
-42	-25	-26	- 93	-70	-38	-108	-69	-27	-10-1	-10-1	-4.2
-40	-24	-24	- 88	-68	-37	-105	-67	-26	- 9-11	- 9-11	-4.1
-37	-22	-22	- 81	-67	-36	-103	-66	-25	- 9-10	- 9-10	-4.0
-35	-21	-20	- 76	-65	-35	-100	-64	-24	- 9-9	- 9-9	-3.9
-34	-20	-19	- 73	-63	-33	- 96	-61	-23	- 9-8	- 9-8	-3.8
-33	-18	-18	- 69	-60	-32	- 92	-59	-22	- 9-6	- 9-6	-3.7
-32	-17	-17	- 66	-57	-30	- 87	-57	-21	- 9-5	- 9-5	-3.6
-30	-16	-16	- 62	-55	-29	- 84	-54	-20	- 9-4	- 9-4	-3.5
-28	-16	-15	- 59	-52	-27	- 79	-52	-19	- 9-2	- 9-2	-3.4
-26	-15	-14	- 55	-49	-26	- 75	-50	-18	- 9-1	- 9-1	-3.3
-25	-14	-13	- 52	-46	-24	- 70	-48	-17	- 9-0	- 9-0	-3.3
-23	-13	-12	- 48	-44	-23	- 67	-45	-16	- 8-10	- 8-10	-3.2
-22	-12	-11	- 45	-41	-21	- 62	-43	-15	- 8-9	- 8-9	-3.1
-20	-12	-10	- 42	-38	-20	- 58	-40	-14	- 8-7	- 8-7	-3.1
-19	-11	- 9	- 39	-35	-18	- 53	-38	-13	- 8-5	- 8-5	-3.0
-18	-10	- 8	- 36	-32	-17	- 49	-35	-12	- 8-4	- 8-4	-2.9
-16	- 9	- 7	- 32	-29	-16	- 45	-33	-11	- 8-2	- 8-2	-2.8
-14	- 8	- 6	- 28	-26	-15	- 41	-31	-10	- 8-0	- 8-0	-2.8
-12	- 7	- 5	- 24	-24	-13	- 37	-29	- 9	- 7-10	- 7-10	-2.7
-10	- 6	- 4	- 20	-22	-12	- 34	-26	- 8	- 7-7	- 7-7	-2.6
- 8	- 5	- 3	- 16	-20	-11	- 31	-23	- 7	- 7-5	- 7-5	-2.6
- 7	- 3	- 2	- 12	-18	-10	- 28	-20	- 6	- 7-2	- 7-2	-2.5

[2] World Book Company.

To read this chart, find the score made by a pupil on any particular subject and find its value in grade or age (both chronological and educational) in the columns to the right. For example, a score of 81 in "Reading Total" is equivalent to an educational age of 9 years, 10 months, and to a beginning fourth-grade standard.

To use the data of the scores of the 5B grade in Woody-McCall Mixed Fundamentals in Arithmetic, Form *I*, given in Table 1, it is necessary to make certain comparisons. To make more meaningful the scores of the class, comparison may be made to the Table of Norms given in Table 5.

TABLE 5.—WOODY-MCCALL STANDARDS FOR THE BEGINNING OF THE SCHOOL YEAR FOR FORM I AND FORM II [3]

Grade	*Standard*	*No. of Cases*
III low	6.1	1,244
III high	10.6	942
IV low	12.5	2,635
IV high	16.4	2,322
V low	17.2	2,395
V high	19.9	2,150
VI low	22.0	2,601
VI high	24.3	2,240
VII low	25.4	2,574
VII high	27.4	1,847
VIII low	27.9	2,060
VIII high	28.5	1,328

To make the Woody-McCall standards comparable with results obtained later in the school year for each month after October, add the following increments:

III	.54
IV	.43
V	.42
VI	.24
VII	.25
VIII	.20

A method of comparing the results of a test with norms is shown in Table 6. The class whose scores are tabulated in this

[3] Bureau of publications, Teachers College, Columbia University.

table took the arithmetic test referred to and should have made a mean score of 17.2, as is shown by the norm given for "V low" in Table 5. The actual achievement of this 5B class is given in Table 1, the mean score being 15.1 and the median score 15. The average of this class is lower than that of "IV high," or 4A, shown in Table 5. The standing of this class is thus roughly determined in relation to the norm. By looking at this table, one can see that this class is approximately one-half year below the norm for its grade.

TABLE 6.—A COMPARISON OF 5B CHILDREN OF CITY SCHOOL X WITH THE CHART NORMS FOR THE WOODY-MCCALL MIXED FUNDAMENTALS IN FORM I, OCTOBER, 1927

	Score	*Total (frequency)*
6B.....	22	1
	21	1
	20	0
	19	1
	18	2
5B.....	17	3
	16	6
Mean of this class.....	15	11
	14	8
	13	3
4B.....	12	0
	11	0
	10	2
	9	1

If it is desirable to find the standing of any particular pupil in comparison with the grade norms, it can easily be done. Take for example the pupil who made a score of 22. This pupil's score is almost a grade ahead of the norm for his grade. That is, he did as well on this test as the average 6B child does

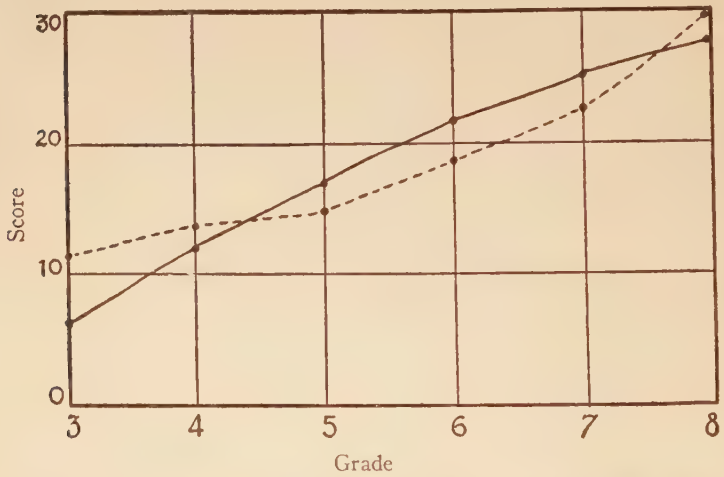


FIG. 5.—A Comparison of Mean Scores of Children in the Various Grades of School X with the Norms for the Woody-McCall Mixed Fundamentals in Form I, October, 1926. The continuous black line represents the norms.

on this test. The pupils that made scores of 9 and 10 were more than a year behind the norm for their class. They made less on this test than the average made by 4B children. If the various grades of this school were compared with the norm from the data in Table 5 the standing of the school as a whole could be determined. The diagram above gives such a comparison.

Figure 6 represents the score of an individual in several subjects as measured by standard tests and a comparison with the norm of his grade in these subjects.

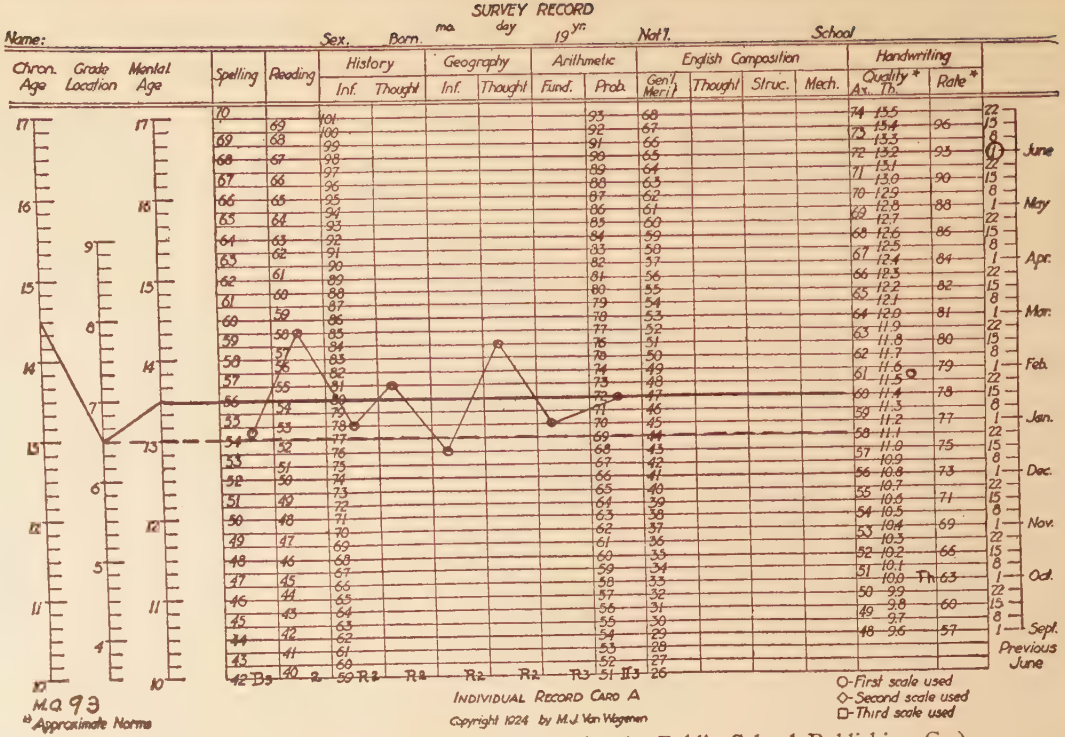


FIG. 6.—Individual Record Card. (Published by the Public School Publishing Co.)

In the preceding figure the dotted line represents the average ability and achievement of pupils who are halfway through the seventh grade, as is this pupil. The heavy straight black line represents the achievement of the average pupil with a mental age of 13-6, or the mental ability of an average child who has lived one-half of his fourteenth year. The thin line connecting the small circles represents the achievement of this particular pupil in standard tests in school subjects. The pupil shown here has a chronological age of 14-6, he is halfway through the seventh grade, and has a mental age of 13-6. His achievement in spelling is about average for his grade. His achievement in reading is much above the average for both his grade and his mental age; on the other hand, his achievement in information geography is distinctly below the average for those of his mental age and lower than the average for his grade.

The Measurement of Relation

Correlation, already mentioned in the discussion of reliability in Chapter III, is a much-used term. Educational literature is replete with references to it. By correlation we mean the extent to which two series of variables tend to change together. If the variables change in the same direction, the correlation is said to be positive. If they change in opposite directions, the correlation is said to be negative. Correlation usually has to do with things that vary in different degrees and is sometimes expressed by a single index of the degree of "going-togetherness" of two variables. This index is the coefficient of correlation, or "*r*," and may be written as $r = -.43$, $r = .84$, etc. The *r* may equal anything from -1.00 to $+1.00$. Perfect negative correlation or opposition is signified by -1.00 and perfect positive correlation by $+1.00$.

Take as an illustration the results obtained when a group of pupils is given a reading test and an arithmetic reasoning test. If those pupils who made high scores in the arithmetic test also made high scores in the reading test, those who made

average or near average scores in one test made like scores in another, and those who made low scores in the arithmetic test also made low scores in the reading test, then it could be said that there was a positive correlation of these two sets of scores. On the other hand, if a pupil's handwriting achievement were measured and his ability in music were measured and these two series of scores were then compared, and if it were found that those pupils who achieved high in handwriting achieved low in music and vice versa, then it would be said that these two measured abilities showed a minus or negative correlation.

Reporting Test Scores

The results of testing are often reported by teachers to their principals or to other teachers in teachers' meetings. Some of these reports are written for professional journals. Administrators report the results of testing also to the community and to their Board of Education. The reporting is made more scientific and vivid by the use of tables and graphs. To make the reporting of the results of measurement by the use of tables accurate, it is necessary to follow certain rules of construction in making these tables. The following rules are listed and illustrated from Table 7.

1. The number and title of the table are placed *above* the table.
2. The title of the table must be complete, inclusive, and clear. See Table 7. It should include such descriptive items as name of test, grades, and date.
3. Other descriptive items should be so placed that they can be easily read. Note: name of schools, grades, etc.
4. Items that are headings to columns should be boxed in. See Grade III, etc., in Table 7.
5. The table reads downward and from left to right. Note: schools in alphabetical order downward and grades from left to right.
6. In tables that need it, rows of dots should be used to guide the eye. Note: Hart School in Table 7.

TABLE 7.—GRADE AND SCHOOL SUMMARY OF ACHIEVEMENT IN ARITHMETIC, AS MEASURED BY THE WOODY-MCCALL MIXED FUNDAMENTALS TEST, FORM I. GRADES 3-8. STAMFORD, CONNECTICUT, DECEMBER, 1922[4]

School	Grade III		Grade IV		Grade V		Grade VI		Grade VII		Grade VIII	
	B	A	B	A	B	A	B	A	B	A	B	A
Center	6.6		13.3	18.3	20.2	21.	21.5	31.	29.3	30.3	30.2	31.2
Cove			17.3	18.7			25.1	26.8				
Elm					21.8	21.	22.5		24.9	25.5		
Franklin			14.9	16.3			28.	27.1				
Glenbrook					16.5	19.3			27.5	26.3		
Hart							22.6	24.			27.5	24.7
Rice	8.7				20.2	22.3			26.4	29.3		
Rogers		13.2					24.3				28.7	32.1
Springdale	6.9		13.3	15.2								
Stevens		8.7			14.	18.1						
Waterside		13.5			19.9	18.5			*	29.		
Willard			19.	20.	24.8	27.2	26.5	29.5	30.	30.5		14.5
Woodland	8.6	9.2										
Entire System	7.8	10.9	14.9	17.1	19.3	19.9	23.5	25.4	27.5	28.6	27.9	25.8

* Due to some errors in giving the test to these half grades the results were not used.

[4] George D. Strayer and N. L. Engelhardt, "Report of the Survey of the Public School System of the Town of Stamford, Conn., School Year, 1922-23," Teachers College, Columbia University.

7. The extra heavy or double line should be used for designation of division of items or thought. Note line between Woodland & Entire System.
8. Important items should be in heavy type.
9. The table should be placed as near as possible to the discussion of it.
10. The reader should be able to place a pencil on any item in the table and readily interpret the meaning of it from the table.
11. Tables may be footnoted. See footnote, Table 7.

The rules of graphing are also listed and partially illustrated.

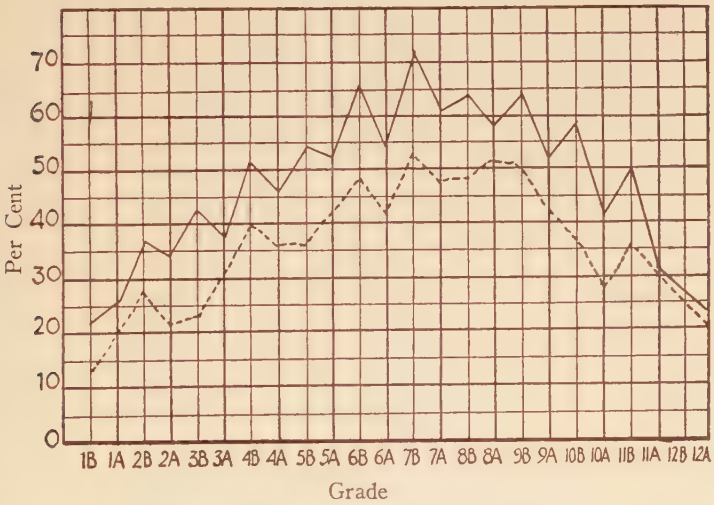


FIG. 7.—A Comparison of Per Cent of Over-ageness by Grades, East Chicago Public Schools, June, 1922, and June, 1926. The broken line represents June, 1926.

1. The title of a diagram should be made as clear and complete as possible. Sub-titles or descriptions should be added if necessary to insure clearness.
2. The general arrangement of a diagram should proceed from left to right.
3. The horizontal scale for curves should usually read from left to right and the vertical scale from bottom to top.

4. Figures for the scales of a diagram should be placed at the left and at the bottom.

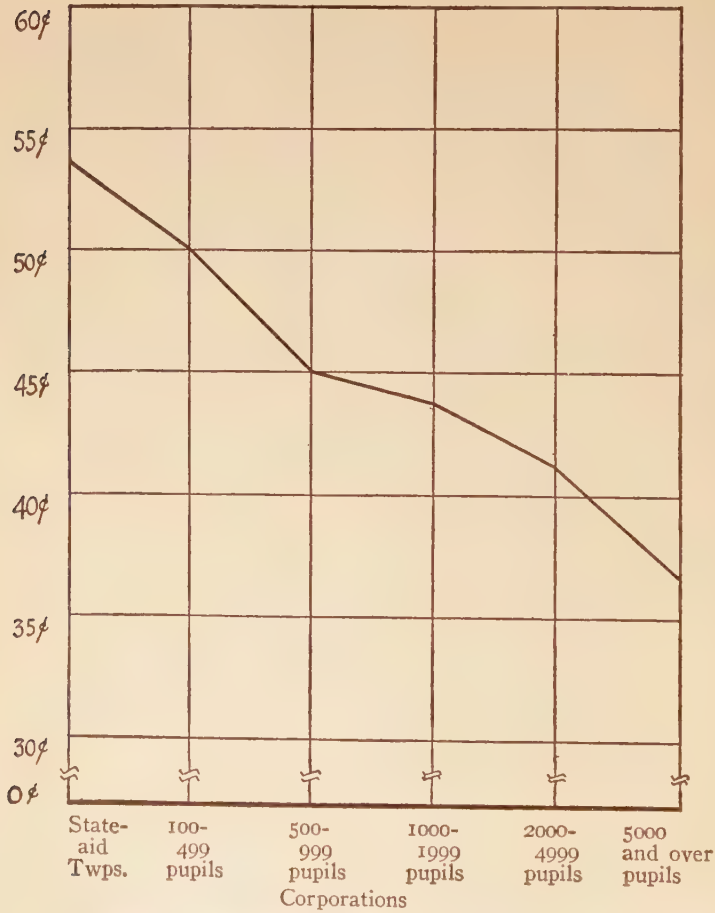


FIG. 8.—Median Price Paid for Chalk (Per Gross) in Indiana School Towns and Cities (Grouped According to School Enrollment) and Also in State-aid Townships.

5. All lettering and all figures on a diagram should be placed so as to be read easily from the base as the bottom.

6. For a curve the vertical scale, whenever practicable, should be so selected that the zero line will appear on the diagram.
7. In graphing two or more bars or curves for comparison, make their zero lines coincide.
8. If the zero line of the vertical scale will not normally appear on the curve diagram, the zero line should be shown by the use of a horizontal break in the diagram. See Figure 8.

When tests are given, they should be given for a definite purpose, and the results should be used for that definite purpose. Whether tests are given for comparison, for diagnosis, for remedial work, or for prognosis, the results should be made intelligibly available to all persons concerned. Certainly, pupils are entitled to know the results of their efforts; otherwise they cannot be expected to correct their defects or to make definite progress. Teachers need to know the results, or they continue in the dark as to the effectiveness of their work and as to the phases of their teaching that need to be improved. Parents are entitled to know not only what their children do but what other children do. Only thus can they be convinced of the degree of efficiency of the schools to which their children go and of the points of strength and weakness in the work of their children. Taxpayers and citizens are entitled to know the results of tests. They are entitled to know how the schools they support rank in comparison with other schools; just where the points of strength are, and where the points of weakness are; and what are the proposed measures for improvement. Where knowledge is, support is apt to be—support in making the poor conditions better and support in keeping the good conditions good.

The methods by which the results of tests are made available to those concerned vary greatly. In general, results are distributed in written reports, or they are given out and explained directly in conferences. Both of these plans should be used. All persons concerned are entitled to receive the

stimulus that comes from the personal presentation of the results by the authority under which the tests were given. They are entitled also to an opportunity for a thorough discussion of the interpretation of the results. Those responsible for giving the tests should therefore meet the children and explain to them the results and the significance of the results. They should meet the teachers and discuss with them the results and the implications of the results. They should meet the parents and citizens generally in appropriate groups where the nature, purpose, and results of the tests, as well as the significance of the findings, could be explained.

Moreover, those concerned are entitled to a permanent record, which can be studied intensively and at leisure, and which will serve as a future reminder and as a constant basis for comparison. Consequently appropriate, tangible, written or printed reports should be put into the hands of pupils, teachers, parents, and citizens generally. For the latter group, articles in the local newspapers would possibly serve best, though occasional printed reports might well be distributed to the homes through the pupils themselves. Periodic printed reports might well be prepared for general distribution in the interest of wholesome and justifiable publicity.

School officials need have no fear in frankly revealing and discussing such results. Any group of people concerned with the welfare of the schools will respond to leadership that has shown the initiative, the daring, and the frankness to reveal and discuss conditions as they are. To such leadership will be gladly entrusted the responsibility for making conditions what they should be.

BIBLIOGRAPHY

Alexander, C. "Presenting Educational Measurements So As to Influence the Public Favorably," *Journal of Educational Research,* Vol 3: pp. 345-58, May, 1921.

Allen, I. M. "A Plan of Reporting the Results of Educational Measurements to Parents," *National Education Association*, Vol. 59: pp. 827-33, 1921.

Barton, W. A. "Pupil Reaction to School Reports," *School Review*, Vol. 33: pp. 77-78, December, 1925; Vol. 34: pp. 42-53, January, 1926.

Brooks, S. S. "The Graphic Interpretation of Test Scores," *Educational Research Bulletin*, Vol. 14, No. 17: p. 366, November, 1925.

Brooks, S. S. "A Universal Graph Record Card for Use with Standard Tests," *Educational Research Bulletin*, Vol. 5, No. 8: pp. 155-61 and 170, April 14, 1926. College of Education, Ohio State University.

Buckingham, B. R. "Suggestions for Procedure Following a Testing Program," *Journal of Educational Research*, Vol. 2: pp. 787-801, December, 1920.

Haggerty, M. E. "Measurement and Diagnosis as Aids to Supervision," *School and Society*, Vol. 6: pp. 271-85, September 8, 1917.

Johnston, Mrs. N. B. "Use of Standardized Educational Tests in School Survey," *Educational Administration and Supervision*, Vol. 11: pp. 588-607. December, 1925.

Madison, I. N. "Procedures Following a Testing Program," *School and Society*, Vol. 13: pp. 600-05, December 24, 1921.

McCall, W. A. and Bixler, H. A. *How to Classify Pupils.* Bureau of Publications, Teachers College, Columbia University, 1928.

Pressey, S. L. "A Simple Apparatus Which Gives Tests and Scores—and Teaches," *School and Society*, Vol. 23: pp. 373-76, March 20, 1926.

CHAPTER V

Arithmetic

The objectives of the teaching of arithmetic most often found in courses of study and literature in the field of curriculum-making are these:[1]

1. The ability and habit of solving arithmetical problems accurately and rapidly.
   a. Ability and habit of performing the operations of addition, subtraction, multiplication, and division with integers, mixed numbers, common and decimal fractions, that are needed in life situations.
   b. Ability and habit of performing the operations with speed and accuracy.
   c. Ability to learn the meaning of numbers and the language of mathematics.
   d. Ability to solve problems: read—pick out important facts—determine process of solution—apply—explain.
2. An understanding of the business transactions of the world and of the quantitative relations needed for insight into the social and economic activities of life. This objective implies that the school should:
   a. Acquaint the child with the business activities of trade, transportation, homes, farm, etc.
   b. Acquaint the child with the arithmetic of civic activities: taxes, school finances, bonded indebtedness, and the quantitative interpretation of civic activities.
   c. Develop within the child an understanding of the advantages of personal thrift, saving and wise spending, and personal accounting and budgeting.

[1] Objectives stated in this chapter as well as in many others are a composite based on the statements of the best authorities in the field. Credit should be given here to Gist, *Elementary School Supervision,* Scribner's; and to Phillips, *Modern Methods and the Elementary Curriculum,* The Century Co.

These objectives, with the study of the activities and abilities necessary to their accomplishment, should form the basis of tests of achievement in the field of arithmetic. The measurement of achievement as well as the diagnosis of difficulties in arithmetic requires some knowledge of the abilities involved in the learning of arithmetic. It is a generally accepted belief that general ability in arithmetic involves the combination of a large number of special skills or abilities in arithmetic. An analysis of the abilities necessary to do the work of addition in the first three grades is found in the following statement by Morton.[2]

The 45 addition combinations were divided into two groups—an easier group of 25 whose sums do not exceed 10 and a harder group of 20 whose sums do exceed 10. In each group the combinations were arranged in approximately the order of their difficulty. . . . It was recognized that each combination, except the doubles, could be arranged in two ways, or that the 45 addition combinations meant 81 facts. The zero combinations were presented later in the kind of addition examples in which zeros would normally occur.

After a few of the combinations had been mastered, the pupil was introduced to easy column addition, three digits in a column. When the 25 easier teaching units had been taught, the 120 three-digit examples whose sums do not exceed 10 were used. After the entire 45 combinations had been taught, the pupil was given practice on 204 additional three-digit examples, examples whose sums exceeded 10 but in which the sum of two digits (adding downward) did not exceed 9.

The next step in the pupil's progress was the learning of higher decade addition, often referred to as "adding by endings." Two types of higher decade addition examples were recognized—those for which the sum was in the same decade as the larger addend, as $14 + 4$, and those which involved "bridging" or in which the sum was in the next higher decade, as $18 + 4$. Specific attention was given to the 225 higher decade addition combinations whose sums do not exceed 39 and

[2] *Teaching Arithmetic in the Intermediate Grades,* p. 13, Silver, Burdett and Company, 1927.

to the 87 additional combinations whose sums exceed 39 and which are needed in carrying on multiplication. Each higher decade combination was specifically practiced in a series of sets of practice examples in column addition.

Carrying was next taught. Two-column examples were limited to six addends and three-column examples were limited to five addends. The pupils were also given practice in adding addends of unequal length and in solving problems involving addition. In all cases the pupils were expected to check their results rather than use answers.

The foregoing quotation, of course, says nothing about the abilities and skills necessary to do division, multiplication, fractions, both common and decimal, and mixed numbers.

In the field of problem solving, four outstanding composite abilities have been pointed out. These are (1) ability to read the problem and find the question, (2) knowledge of what facts can be used to find the answer to the question or problem, (3) knowledge of what process to use in the solution, (4) use of the fundamental processes to solve the problem. The ability to read the problem and find the question, of course, involves many abilities or skills such as understanding of the general vocabulary of the problem and of the special technical vocabulary of arithmetic. By continued analysis, the abilities and skills necessary to do the work outlined by the course of study will eventually be listed. Such an analysis will aid in insuring better teaching and also more valuable testing.

A knowledge of the abilities that are necessary to attain the desirable objectives of arithmetic does not necessarily insure use of the best method of testing. The function of the test itself has much to do with the method of testing. The measurement of achievement in arithmetic computation usually undertaken is measurement of the (1) rate or speed of doing problems that represent certain skills, (2) accuracy with which they are done, (3) power to do increasingly difficult problems, (4) combinations of these three. The first two tend to go together; that is, the same habits which yield accurate results

will usually produce speed. This statement, of course, does not necessarily apply to a particular individual, since a particular pupil may be slow but accurate. The third type of measurement is measurement of the power to do increasingly difficult problems. The use of this measurement of arithmetic ability has been criticized on the ground that social utility and not difficulty is the basis for determining the content of the curriculum and that social utility, therefore, should be measured in testing. It is pretty generally admitted that the computation difficulties of such things as cube root and greatest common divisor are poor measures of the results of teaching from present-day textbooks.

Most of the tests of the general type include (1) a measure of speed by timing; (2) a measure of accuracy by requiring that the answer be correct if credit is given; (3) a gradation of problems beginning with those that are easier and continuing with those of increased difficulty.

Diagnostic tests in arithmetic attempt, on the other hand, to cover a very wide field of specific abilities and usually include a test of those skills that are needed in the solution of the exercises or problems. The more specific diagnosis, of course, requires an individual step-by-step working out of the problem by the pupil. Usually the time element is not included in diagnostic tests. This may be a mistake because, as a matter of fact, it is desirable to know the difficulties of pupils in their life situation; these situations are often, in a sense, timed. If diagnostic tests are to represent typical pupil difficulties, they might also be timed.

Reasoning problem tests are on the whole quite different from the regular computation type. They are usually of the fourth type, i.e., timed, scored on accuracy of answer, and made up of problems graduated as to difficulty. Some attempts have been made recently to develop specific measures of certain phases of problem solving by testing the ability of the pupil to analyze a problem. These attempts are somewhat diagnostic.

There are several general survey tests that are valuable in providing a measure of the pupil's ability to perform the operations of arithmetic. Some of these tests are partly diagnostic. With the development of truly diagnostic tests, however, these tests will serve mostly as general survey tests. Such tests as the Monroe Standardized General Survey Arithmetic Scales, Courtis Standard Research Tests, Series *B,* Compass Survey Test, and the Woody-McCall Mixed Fundamentals are examples of this general survey type. These tests measure in a selective fashion the first stated objective—"The ability and habit of solving arithmetical problems accurately and rapidly."

The Monroe Standardized General Survey Arithmetic Scales [3] are distinctly a survey test and offer little opportunity for diagnosis. Scale I of this test is for the third, fourth, and fifth grades; Scale II is for grades six, seven, and eight. Equivalent forms are available for each scale. Each scale is spiral in type, and the various parts are given as single tests which allow for separate scores for each part and a single score for the whole test. Norms in terms of grades and ages are available, although they are not highly standardized. The reliability of the test is indicated by a coefficient of correlation of reliability, which is .76. Its validity is only fairly high. Part of the test is given here.

Test No. 1—ADDITION

No. Right.....

7862	6809	8941	5917	6772	7864	1249
5013	7623	7910	4814	6028	7883	8975
1761	5299	9845	9007	6535	8240	9005
5872	6601	8522	6975	2340	9869	1573
3739	3496	1046	1227	2319	6794	3203

[3] Public School Publishing Co. This test is a part of the composite Illinois Examination, and is also published separately.

Test No. 2—MULTIPLICATION

No. Right.....

4857	5718	6942	4065
36	92	58	47

Test No. 5—ADDITION AND SUBTRACTION OF FRACTIONS

No. Right..... × 2 = ...

$\frac{1}{6} + \frac{1}{3} =$ $\frac{3}{4} - \frac{2}{5} =$ $\frac{1}{6} + \frac{3}{5} =$

The test may be criticized adversely on the grounds that its usefulness is limited to survey purposes and that it measures relatively few abilities in the field of arithmetic computation. The giving of the parts requires rigid adherence to a very short time limit. A stop watch and a careful tester are necessary to its use if consistent results are expected.

The Stanford Arithmetic Examination may be used as a survey test of ability in performing operations and solving problems. This test is outstanding from many points of view and is here described in some detail.[4] This test is made up of two parts, computation and reasoning. Each yields a separate score, but the two are combined for the total arithmetic score.

The computation test consists of 47 examples selected from leading textbooks through the judgments of the authors of the test and by experimental try-outs. These examples are of increasing difficulty but include mostly examples of a type taught in the elementary school. A wide range of abilities is measured. The coefficient of reliability for grades and ages is given in Table 8. The validity of the Stanford Arithmetic Examination as measured by correlation with other tests and teachers'

[4] The material of these tests is used for both primary grades (2 and 3) in a shorter form, and advanced grades (4-8) in the form here described. Published by World Book Company.

marks is higher than for most other tests in arithmetic. Age and grade norms are available.

The reasoning test consists of 40 problems chosen somewhat in the same way as the computation examples. The problems are of increasing difficulty. These problems minimize the computation and emphasize the reasoning side of arithmetic. Most of the problems meet the demands of social utility. The coefficient of reliability is given in Table 8. Equivalent forms are available.

TABLE 8.—RELIABILITY COEFFICIENTS OF THE ARITHMETIC TESTS BY AGES AND GRADES [5] (STANFORD ACHIEVEMENT)

BY AGES			BY GRADES		
Age	*Arith. Comp.*	*Arith. Reas.*	*Grade*	*Arith. Comp.*	*Arith. Reas.*
7	.87	.87	2	.74	.83
8	.90	.84	3	.80	.72
9	.92	.87	4	.75	.67
10	.92	.88	5	.81	.88
11	.89	.87	6	.77	.71
12	.85	.90	7	.65	.79
13	.85	.91	8	.69	.70
14	.83	.88	9	.70	.85
15	.76	.89			

The preceding table may be read as follows:

If we select seven-year-old pupils, the correlating of scores made by these pupils on Form *A* of the test with scores made on Form *B* of the test should give a correlation of .87 in Arithmetic Computation; in Arithmetic Reasoning, .87. If a group of second grade pupils were selected and tested in the same manner, the correlation for Arithmetic Computation would be .74 and Arithmetic Reasoning .83.

The scoring of the test is highly objective. The manual of directions for giving and scoring is clear, specific, and definite. The scoring keys are easy to use. Age and grade norms are available.

[5] Taken from the manual for the Stanford Achievement Test, Advanced Examination. World Book Company.

This test has been widely and successfully used both separately and as a part of the composite test, the Stanford Achievement Test.

Following are the directions to be used in giving the Stanford Arithmetic Examination. Samples of the norms and of the directions from the scoring key are also given.

PROCEDURE, ADVANCED EXAMINATION, GRADES 4 TO 8

(Identical for Form *A* and Form *B*. To be followed verbatim)

"Here is a test to show how much you have learned. I will give each of you a test book. Do not write on it or open it until I tell you to." (See that this is obeyed.)

After all are provided with test books and pencils: "Now fill the blanks at the top of the first page, here." (Hold up test book and point to the blanks.) "Do it as quickly as you can but write plainly. On the first line where it says 'Name,' write your name." (Pause.) "After the word 'Grade,' write the number that tells what grade you are in." (Name the grade.) "After 'Boy or Girl,' write the word that tells which you are." (Pause.) "On the next line, where it says 'Age,' tell how old you are now." (Pause.) "Then tell when your next birthday will come and how old you will be then." (Pause.) "On the next line write the name of this school." (Give the name.) "At the end of the same line give the date." (Name date. Give to the younger children any necessary assistance.)

After the blanks have been filled: "Now, listen carefully and do just what I tell you to do. Do not begin until I say 'Go.' The very second I say 'Stop,' you must stop and hold your pencils up. After we have begun, you must not ask questions. If you break your pencil, hold up your hand and I will give you another. Do your best and do not pay any attention to what any one else is doing. We want to see if you have learned as much as the children in other schools.

Test 4. Arithmetic: Computation

Give each pupil two or three sheets of blank paper to figure on.

"Turn to Test 4." (Be sure that all have found the place.)

"Read the directions at the top of the page: 'Get the answers to these examples as quickly as you can without making mistakes. Look carefully at each example to see what you are to do.'" (Pause slightly.) "You may use the blank sheets of paper on your desk to figure on if you need to. There are two pages in this test. As soon as you have finished the first page, go right on to the next. Ready—Go." (See that pupils do not stop at end of first page.)

Allow *20 minutes;* then say: "Stop. Turn to Test 5 on the next page. Be sure to turn just one leaf." (Pause until all have found the place.)

Test 5. Arithmetic: Reasoning

See that each pupil has enough blank paper to figure on.

"Read the directions at the top of the page: 'Find all the answers as quickly as you can. Write the answers on the dotted lines. Use the blank sheets of paper to figure on.'" (Pause slightly.) "The test has two pages. As soon as you have finished the first page, go right on to the next. Ready—Go." (See that pupils do not stop at end of first page.)

Allow *20 minutes;* then say: "Stop. Close your books." (Second sitting ends here.)

Test 4. Arithmetic: Computation [6]

Get the answers to these examples as quickly as you can without making mistakes. Look carefully at each example to see what you are to do.

Begin here

(1)	(3) Add	(4) Add	(5) Add
$3+2=$	2 5	7 4	13 2

(21) Add	(22) Multiply	(23) Multiply	(24)	(25)
24 $12\frac{4}{5}$	6389 7	4679 68	$2)\overline{15.8}$	$2\frac{7}{8}-1=$

[6] From Arithmetic Examination, Form *A*, for Grades 2-8, of the Stanford Achievement Test, by Truman L. Kelley, Giles M. Ruch, and Lewis M. Terman, World Book Company.

(38)
$27.34 + 2\frac{1}{4} + 89.2 + 4\frac{3}{4} =$

(39)
$3\frac{1}{4} \times 5\frac{1}{2} \times 3\frac{1}{2} =$

(40)
$1\frac{3}{4} + 25.2 + 4\frac{1}{5} + 48.961 =$

Test 5. Arithmetic: Reasoning

Find all the answers as quickly as you can. Write the answers on the dotted lines. Use the blank sheets of paper to figure on.

Begin here

1 How many are 3 eggs and 2 eggs? *Answer*.....
2 Mary is 7 years old. How old will she be in 3 years? *Answer*.....
3 A hen had 9 chicks and 3 of them died. How many were left? *Answer*.....
26 If a fence rail is 10 feet long, how many rails will it take to reach a mile? *Answer*.....
27 Sound travels about 1100 ft. a second. If you see the flash of a cannon and 12 seconds later the sound reaches you, how far away is the cannon? *Answer*.....
28 A man had $5000, from which he received 6 per cent income each year. In addition he earned $1500 in business. What was his total income for the year? *Answer*.....
29 Frank and George buy 300 marbles for 50 cents. Frank pays 35 cents and George 15 cents. How many marbles should George receive? *Answer*.....
35 A man invested $1000 in each of 3 different bonds. The first paid 8 per cent dividend and the second 6 per cent, but on the third he lost $5 on each hundred dollars invested. What was his net yearly gain on the three investments? *Answer*.....
36 If the circumference of a circle is 12.5664 feet, what is its diameter? *Answer*.....
37 The regular price of a certain piece of linen is $4 per yard. A remnant $1\frac{1}{4}$ yards long is offered at $2.50. What per cent reduction is made? *Answer*.....

DIRECTIONS FOR SCORING ADVANCED EXAMINATION, FORM *A*

The following general rules apply in the scoring of all the tests:

1. Items attempted are scored either right or wrong. No part credits are given.

2. When an item evidently has been corrected by the pupil, the correction is the answer which should be scored.

3. The score for each test should be entered in the space provided in the lower right-hand corner of the test page.

Test 4. Arithmetic: Computation

1. Score is number right multiplied by 4.

2. If two answers are given to a problem, count as wrong.

3. An answer may be credited even if somewhat misplaced, provided it is clearly intended as the answer to the problem in question.

4. Omission of dollar sign or words "cents," "feet," etc., is permissible.

5. Fractions expressed as decimals receive full credit.

6. An answer which is a common fraction must be reduced to lowest terms to receive credit.

Test 5. Arithmetic: Reasoning

1. Score is number right multiplied by 4.

2. Answer may be written on the dotted line or elsewhere near its problem.

3. If two different answers are given to any problem, count as wrong.

4. Omission of dollar sign or words "cents," "feet," etc., is permissible.

5. Omission of decimal point is permissible in Form *A*, Items 14, 16, and 23.

6. Fractions expressed as decimals receive full credit.

7. An answer which is a common fraction must be reduced to lowest terms to receive credit.

Key for Advanced Examination, Form A, Test 4

(1)	5	(6)	19
(2)	7	(7)	2
(3)	7	(8)	3
(4)	11	(9)	6
(5)	15	(10)	101

TABLE 9.—STANFORD ACHIEVEMENT TEST. EQUIVALENT SCORES FOR ARITHMETIC TEST. ARITHMETIC COMPUTATION, ARITHMETIC REASONING, AND TOTAL

Arith. Comp.	*Arith. Reas.*	*Arith. Total*	*Subject Age*	*Chronological Age*	*Grade*
– 40	**– 20**	**– 60**	**– 8-9**	**– 8-9**	**– 3.1**
– 42	– 22	– 64	– 8-10	– 8-10	– 3.1
– 45	– 23	– 68	– 8-11	– 8-11	– 3.2
– 47	– 25	– 72	– 9-0	– 9-0	– 3.3
– 50	– 26	– 76	– 9-1	– 9-1	– 3.4
– 53	**– 27**	**– 80**	**– 9-2**	**– 9-2**	**– 3.4**
– 54	– 29	– 83	– 9-4	– 9-4	– 3.5
– 56	– 30	– 86	– 9-5	– 9-5	– 3.6
– 57	– 30	– 87	– 9-6	– 9-6	– 3.7
– 59	– 31	– 90	– 9-7	– 9-7	– 3.8
– 60	**– 32**	**– 92**	**– 9-8**	**– 9-8**	**– 3.9**
– 62	– 33	– 95	– 9-10	– 9-10	– 4.0
– 63	– 34	– 97	– 9-11	– 9-11	– 4.0
– 65	– 35	–100	–10-0	–10-0	– 4.1
– 66	– 36	–102	–10-1	–10-1	– 4.2
– 68	**– 37**	**–105**	**–10-2**	**–10-2**	**– 4.3**
– 70	– 38	–108	–10-4	–10-4	– 4.4
– 71	– 38	–109	–10-5	–10-5	– 4.4
– 73	– 39	–112	–10-6	–10-6	– 4.5
– 74	– 41	–115	–10-7	–10-7	– 4.6
– 76	**– 42**	**–118**	**–10-8**	**–10-8**	**– 4.7**
– 78	– 43	–121	–10-9	–10-9	– 4.8
– 80	– 44	–124	–10-9	–10-9	– 4.9
– 82	– 46	–128	–10-10	–10-10	– 4.9
– 84	– 47	–131	–10-11	–10-11	– 5.0
– 86	**– 48**	**–134**	**–11-0**	**–11-0**	**– 5.1**
– 87	– 50	–137	–11-1	–11-1	– 5.2
– 89	– 51	–140	–11-2	–11-2	– 5.3
– 91	– 52	–143	–11-3	–11-3	– 5.4
– 93	– 54	–147	–11-4	–11-4	– 5.5
– 95	**– 55**	**–150**	**–11-5**	**–11-5**	**– 5.6**
– 96	– 56	–152	–11-6	–11-6	– 5.7
– 98	– 58	–156	–11-7	–11-7	– 5.7
–100	– 59	–159	–11-8	–11-8	– 5.8
–102	– 60	–162	–11-9	–11-9	– 5.9
–103	**– 61**	**–164**	**–11-10**	**–11-10**	**– 6.0**
–105	– 62	–167	–11-11	–11-11	– 6.1
–107	– 63	–170	–12-0	–12-0	– 6.2
–109	– 65	–174	–12-1	–12-1	– 6.3
–110	– 66	–176	–12-2	–12-2	– 6.4
–112	**– 67**	**–179**	**–12-4**	**–12-4**	**– 6.5**
–113	– 68	–181	–12-5	–12-5	– 6.6
–114	– 70	–184	–12-6	–12-6	– 6.7
–115	– 72	–187	–12-7	–12-7	– 6.8

TABLE 9.—STANFORD ACHIEVEMENT TEST. EQUIVALENT SCORES FOR ARITHMETIC TEST. ARITHMETIC COMPUTATION, ARITHMETIC REASONING, AND TOTAL

Arith. Comp.	*Arith. Reas.*	*Arith. Total*	*Subject Age*	*Chronological Age*	*Grade*
–116	– 73	–189	–12-8	–12-8	– 7.0
–117	– 75	–192	–12-9	–12-9	– 7.1
–117	– 77	–194	–12-11	–12-11	– 7.2
–118	– 78	–196	–13-0	–13-0	– 7.3
–119	– 80	–199	–13-2	–13-2	– 7.4
–120	– 81	–201	–13-3	–13-3	– 7.5
–121	– 83	–204	–13-5	–13-5	– 7.6
–123	– 84	–207	–13-6	–13-6	– 7.7
–124	– 86	–210	–13-8	–13-8	– 7.9
–125	– 87	–212	–13-10	–13-10	– 8.0
–127	– 89	–216	–14-0	–14-0	– 8.1
–128	– 91	–219	–14-2	–14-2	– 8.2
–129	– 92	–221	–14-4	–14-4	– 8.4
–131	– 93	–224	–14-6	–14-6	– 8.5
–132	– 95	–227	–14-8	–14-9	– 8.6
–133	– 96	–229	–14-9	–15-0	– 8.8
–134	– 97	–231	–14-11	–15-4	– 8.9
–136	– 98	–234	–15-1	–15-9	– 9.0
–137	– 99	–236	–15-2	–16-2	– 9.2
–138	–100	–238	–15-4	–16-9	– 9.3
–139	–101	–240	–15-6	–17-9	– 9.5
–140	–102	–242	–15-7	–Adult	– 9.7
–141	–103	–244	–15-9		– 9.8
–142	–104	–246	–15-10		–10.0

To use this table of equivalent scores or norms, (1) find in the column under Arithmetic Computation the score made by a pupil; (2) look under column marked Subject Age directly opposite the score made by the pupil—this is his subject age. In the last column directly opposite his subject age is the grade equivalent of the same score. For example, a pupil makes a score of 112 in Arithmetic Computation, which is equivalent to a subject age of 12 years and 4 months. This means that the score 112 is the average score made by pupils aged 12 years and 4 months. This is equivalent in grade to 6.5 or to the average score made by pupils halfway through the sixth grade.

The Stanford Arithmetic Examination yields a subject quo-

tient. This is derived by dividing the pupil's subject age (which is the average age of those pupils making the particular score) by the actual chronological age of the pupil. Suppose that a score of 184 on the arithmetic test (total) is made by pupil M, who is 12 years of age. By looking at the equivalent age scores shown in Table 9, it is seen that 184 is equivalent to 12 years and 6 months. By changing this age to months and substituting in the formula, the quotient is obtained, thus:

$$\frac{\text{Subject age (in months)} \times 100}{\text{Actual age (in months)}} = \text{Arithmetic quotient}$$

$$\frac{150 \times 100}{144} = 104+$$

The arithmetic quotient is 104. This means that pupil M achieved better than the average child of his age. This fact is represented by the single index of arithmetic quotient of 104. If pupil X has an arithmetic quotient of 93, it means that he is achieving less as measured by this test than are others of his own age. Such a quotient may easily point to the need of special diagnostic testing. Such testing should be for the purpose of revealing the particular difficulty of a child whose arithmetic quotient is far below 100. It may also point to need of extra promotion, reclassification, or enriched curriculum for pupils with an AQ (arithmetic quotient) of 120 or above.

The achievement of a class, a school, or a city can be made meaningful by comparisons in terms of age and grade norms. This is shown by the diagram, Figure 5, Chapter IV.

While the Stanford Arithmetic Examination makes possible some diagnosis, it is necessarily fragmentary in this respect. It can therefore be more profitably used to point out pupils who are deficient in arithmetic and to give some clue as to the phase of the test that any particular pupil finds difficult. This test should then be followed by a more diagnostic type of test.

DIAGNOSTIC TESTS

Disability in arithmetic may be indicated by a survey test. However, disability must be further analyzed, first to find the type of errors and then to find the cause of these errors. When this analysis can be made in terms of teaching units and methods, it is desirable to have it so made. The remedial work is then more easily carried out. A truly diagnostic test would measure accurately all the separate arithmetic skills of any social value, but it would be entirely too long to be used. In a short test the more nearly these skills are covered by intelligent and scientific sampling, the more truly will the test be diagnostic. Such a test should make it possible to test any special phase of the arithmetic work and thus reveal a picture of the specific weakness of the pupil. Such results ought to lead to a program of remedial instruction. It should be possible to re-test with equivalent forms to determine progress in the remedial work.

An outstanding example of this kind of test is the Compass Diagnostic Tests in Arithmetic.[7] A detailed statement of the content of these tests follows:

THE SCOPE OF THE COMPLETE TESTING PROGRAM

Below is a catalog of the twenty tests, part by part, showing the extent of the diagnostic testing possible. The grades for which each test was designed and standardized are also indicated.

TEST I: ADDITION OF WHOLE NUMBERS (Grades 2-8)
- Part 1, *Basic Addition Facts*
- Part 2, *Higher Decade Addition*
- Part 3, *Column Addition*
- Part 4, *Carrying in Column Addition*
- Part 5, *Checking Answers in Addition*

[7] Scott, Foresman and Company.

TEST II: SUBTRACTION OF WHOLE NUMBERS (Grades 2-8)
- Part 1, *Basic Subtraction Facts*
- Part 2, *Harder Subtraction*
- Part 3, *Checking Answers in Subtraction*
- Part 4, *Finding Errors in Subtraction*

TEST III: MULTIPLICATION OF WHOLE NUMBERS (Grades 3-8)
- Part 1, *Basic Multiplication Facts*
- Part 2, *Additions Used in the Multiplication of Parts 4, 5, and 6*
- Part 3, *Carrying in Addition Used in the Multiplication of Parts 4, 5, and 6*
- Part 4, *Fundamentals in Multiplication*
- Part 5, *Checking Multiplication*
- Part 6, *Finding Errors in Multiplication*

TEST IV: DIVISION OF WHOLE NUMBERS (Grades 4-8)
- Part 1, *The Vocabulary of Division*
- Part 2, *Fundamentals of Short Division*
- Part 3, *Short Division with Carrying*
- Part 4, *Multiplication, Addition, and Subtraction Used in Division in Parts 2, 3, 5, 6, and 7.*
- Part 5, *Estimating the First Quotient Figure*
- Part 6, *Fundamentals of Long Division; Checking*
- Part 7, *Finding Errors in Long Division*

TEST V: ADDITION OF FRACTIONS AND MIXED NUMBERS (Grades 5-8)
- Part 1, *Changing Fractions to Equivalent Forms*
- Part 2, *Finding Common Denominators*
- Part 3, *Reducing Fractions to Lowest Terms*
- Part 4, *Fundamentals of Addition of Fractions*
- Part 5, *Reducing Answers and Finding Errors*

TEST VI: SUBTRACTION OF FRACTIONS AND MIXED NUMBERS (Grades 5-8)
- Part 1, *Changing Fractions to Equivalent Forms*
- Part 2, *Finding Common Denominators*
- Part 3, *Expressing Mixed Numbers as Improper Fractions*
- Part 4, *Fundamentals of Subtraction of Fractions*
- Part 5, *Reducing Answers and Finding Errors*

TEST VII: MULTIPLICATION OF FRACTIONS AND MIXED NUMBERS (Grades 5-8)

Part 1, *Changing Mixed Numbers to Improper Fractions*
Part 2, *Cancellation in Multiplication of Fractions*
Part 3, *Reducing Answers to Best Form*
Part 4, *Fundamentals of Multiplication of Fractions*
Part 5, *Finding Errors*

TEST VIII: DIVISION OF FRACTIONS AND MIXED NUMBERS (Grades 5-8)

Part 1, *Changing Mixed Numbers to Improper Fractions*
Part 2, *Cancellation in Division of Fractions*
Part 3, *Reducing Answers to Best Form*
Part 4, *Changing from Multiplication to Division Form*
Part 5, *Fundamentals of Division of Fractions*
Part 6, *Finding Errors*

TEST IX: ADDITION, SUBTRACTION, AND MULTIPLICATION OF DECIMALS (Grades 5-8)

Part 1, *Vocabulary and Notation of Decimals*
Part 2, *Changing Fractions and Mixed Numbers to Decimal Form, and the Reverse*
Part 3, *Fundamentals of Addition and Subtraction of Decimals*
Part 4, *Place Value and Pointing in Multiplication of Decimals*
Part 5, *Finding Errors in Pointing Answers*

TEST X: DIVISION OF DECIMALS (Grades 6-8)

Part 1, *Dividing Decimals by Pointing*
Part 2, *Locating Decimal Points in Division*
Part 3, *Changing Remainders to Decimal Form*
Part 4, *Fundamentals of Division of Decimals*
Part 5, *Finding Errors*

TEST XI: ADDITION AND SUBTRACTION OF DENOMINATE NUMBERS (Grades 6-8)

Part 1, *Knowledge of Tables of Measure*
Part 2, *Reducing and Borrowing in Denominate Numbers*
Part 3, *Fundamentals of Addition of Denominate Numbers*

Part 4, *Fundamentals of Subtraction of Denominate Numbers*
Part 5, *Finding Errors in Addition and Subtraction of Denominate Numbers*

TEST XII: MULTIPLICATION AND DIVISION OF DENOMINATE NUMBERS (Grades 6-8)
Part 1, *Knowledge of Tables of Measure*
Part 2, *Reducing Denominate Numbers to Larger or Smaller Units*
Part 3, *Fundamentals of Multiplication of Denominate Numbers*
Part 4, *Fundamentals of Division of Denominate Numbers*
Part 5, *Finding Errors*

TEST XIII: MENSURATION (Grades 7-8)
Part 1, *Vocabulary of Mensuration*
Part 2, *Mensuration of Plain Surfaces*
Part 3, *Mensuration of Solids*
Part 4, *Finding Areas and Volumes*
Part 5, *Writing Formulas Used in Mensuration*

TEST XIV: THE BASIC FACTS OF PERCENTAGE (Grades 6-8)
Part 1, *Equivalent Fraction, Decimal, and Per Cent Relations*
Part 2, *Expressing Areas in Per Cents*
Part 3, *Choosing Correct Solutions in Percentage*
Part 4, *Easy Work in Percentage*
Part 5, *Harder Work in Percentage*

TEST XV: INTEREST AND BUSINESS FORMS (Grades 7-8)
Part 1, *Vocabulary of Interest and Business Forms*
Part 2, *Common Business Forms and Budgets*
Part 3, *Computation of Interest, Discount, Etc.*
Part 4, *Use of Interest Tables*

TEST XVI: DEFINITIONS, RULES, AND VOCABULARY OF ARITHMETIC (Grades 4-8)
Part 1, *Basic Vocabulary and Definitions of Arithmetic*
Part 2, *Basic Rules of Arithmetic*

TEST XVII: PROBLEM ANALYSIS: ELEMENTARY (Grades 5-6)
Part 1, *Comprehension*
Part 2, *What Is Given*
Part 3, *What Is Called For*
Part 4, *Probable Answer*
Part 5, *Correct Solution*

TEST XVIII: PROBLEM ANALYSIS: ADVANCED (Grades 7-8)
Part 1, *Comprehension*
Part 2, *What Is Given*
Part 3, *What Is Called For*
Part 4, *Probable Answer*
Part 5, *Correct Solution*

TEST XIX: GENERAL PROBLEM SCALE: ELEMENTARY (Grades 5-6)

TEST XX: GENERAL PROBLEM SCALE: ADVANCED (Grades 7-8)

To give some idea of the foregoing tests and their use, parts of the tests are here reproduced and discussed.

Test I. Addition of Whole Numbers: Form *A*, primarily for grades two and three although it can be used also in higher grades, is made up of five parts. Part 1 is a test on the basic addition facts; seventy of the 100 basic combinations are included in the test. Checking them against such a study of difficulty of combinations as that made by Clapp,[8] shows that the seventy combinations included are well distributed. Also, both $8 + 5$ and $\begin{array}{r} 7 \\ +\ 4 \\ \hline \end{array}$ types of addition are included. There are 28 examples of the first type and 42 of the second type.

Part 2, Higher Decade Addition, has 66 examples out of a possible more than 300. These are well selected and include

[8] *The Number Combinations: Their Relative Difficulty and Frequency of Their Appearance in Textbooks.* Bureau of Educational Research Bulletin No. 1 and 2. University of Wisconsin, Madison, 1924.

approximately one-third that require the bridging of the ten. The totals are all under 50. The positions of the larger and smaller numbers are interchanged, and both column and + type of addition problems are included.

Part 3, Column Addition, consists of 20 examples of one-digit column addition. The problems are of the 3, 4, 5, 6, and 7 number type. The totals are all less than 50.

Part 4, Carrying in Column Addition, contains thirteen examples, which range from examples of two digits with single carrying to examples of four digits, seven high, and three carryings. Eleven examples are given in column form and two in line form which requires the pupil to copy before adding.

Part 5, Checking Answers in Addition, requires the pupil to check the example by adding down.

Although each test yields a separate weighted score, these are all brought together in a single total score. Mid-year and end-year norms are available for each test by parts and by total. Age equivalents or norms are available for total score on each test. Since the tests are timed and provide for a single score, the results might be of some value as a survey test. However, the time limits are rather long. In some grades pupils finish before the end of the time. The Compass Diagnostic Tests may be said to measure the objectives of arithmetic stated at the beginning of this chapter more nearly than does any other series of tests that has as yet been developed.

The manual of directions is complete, and the directions for giving and scoring are clear, comprehensive, and definite. The scoring is objective. The mechanics of the scoring key could be improved. An equivalent form is not yet available but is in preparation. The test might be improved by better numbering of examples and by the inclusion of more detailed instructions in a few places. The timing must be carefully done if the results are to be used in any sense for survey purposes.

The directions for giving follow:

PRELIMINARY INSTRUCTIONS TO THE PUPILS: ALL TESTS

After the booklets or test sheets have been distributed to the pupils in the manner described above, say to the pupils: "Turn over the test sheet (or booklet) to the front page. Write in your name, grade, age, and the other facts called for. Write your name here." (Point to place.) "Write quickly but plainly." (Pause.) "Now write the grade you are in here." (Point to place.)

The examiner should continue for the other pupil-information blanks in the same way, pointing out the spaces and pausing to allow the pupils time to fill the blanks. In the upper grades the pupils can be depended on to proceed without help, but in the lower grades the examiner should write such facts as the name of the school on the blackboard. With the very youngest children, it will usually be necessary for the teacher to supply most of the information from her school register.

Care must be taken to see that the information about age is written in such a way that the exact year and month of age may be determined. For this reason the blanks provide a double check by asking for "Age," "When is your next birthday?" and "How old will you be then?"

After the blanks have been filled in, say: "Now listen carefully and be sure to do exactly what I tell you to do. Do not begin until I say 'Go.' When I say 'Stop,' you must stop work at once. If you break your pencil, hold up your hand and I will give you another. Do you understand?" (Explain any misunderstandings at once.)

NOTE: *Follow at once with the specific directions for the test you are giving.* These are printed on the pages immediately following.

DIRECTIONS FOR TEST I: ADDITION OF WHOLE NUMBERS

Say: "This is a test to see how well you can add numbers. We want to see if you can do as well as pupils in other schools. Be sure to do the very best you can.

"Look at the direction for Part 1." (Point to place.) "It says 'Add.' You are to find the sums for the examples as quickly as you can without making mistakes. Begin here" (point) "and work down the two columns and then across the

page this way." (Demonstrate.) "Stop at the end of Part 1 where the line goes across the page." (Point out the line which separates Parts 1 and 2.) "Ready; go!" (Record the time.)

Allow 3 minutes; then say: "Stop. Look at the direction for Part 2 right below where you have been working." (Point.) "You are to add just as you did the others. Remember, after you finish the work on this page, to turn over and finish the two rows of examples at the top of the next page. Ready; go!"

Allow 6 minutes; then say: "Stop. Look at Part 3. It says 'Add.' Find the sums for the twelve examples. Ready; go!"

Allow 3 minutes; then say: "Stop. Look at Part 4. It says 'Add.' Find the sums for the thirteen examples. Notice that the last two must be copied at the right where the numbers 12 and 13 stand. Ready; go!"

Allow 6 minutes; then say: "Stop. Look at Part 5. It says: 'Directions: Some of the printed answers below are right, and some are wrong. Check each sum by adding *downwards*. Write your check answer on the line at the *top* of the example. The first one is already done correctly.' Ready; go!"

Allow 9 minutes; then say: "Stop. Turn over your test sheet to the first page. See if you have forgotten to fill in any of the blanks."

Collect the tests at once, making sure that all the pupil-information blanks have been correctly filled.

TEST 1: ADDITION OF WHOLE NUMBERS: FORM A[9]

Part 1—Basic Addition Facts

Add:														
0 + 2 =	0	9	5	2	3	1	7	9	8	5	7	6	4	2
1 + 7 =	2	1	0	3	4	1	5	2	4	4	4	2	1	6
2 + 2 =	—	—	—	—	—	—	—	—	—	—	—	—	—	—
7 + 6 =														
2 + 5 =														
2 + 1 =	9	4	8	0	4	2	6	5	2	4	3	7	8	8
0 + 9 =	5	0	6	7	2	5	6	3	4	5	7	3	3	8
9 + 2 =	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3 + 9 =														
5 + 8 =														
9 + 1 =														
8 + 5 =														
2 + 8 =														

[9] From the Compass Diagnostic Tests in Arithmetic, by Ruch, Knight, Greene, and Studebaker. Published by Scott, Foresman and Company.

Part 2—Higher Decade Addition

Add:

2	1	6	8	18	4	9	7	24	10	6	16
20	21	22	13	6	12	29	25	7	2	12	6

19 + 7 =
12 + 2 =

14 + 2 =
11 + 7 =

Part 3—Column Addition

Add:

4	7	8	2
7	4	0	4
5	9	1	3
	1	2	8
		4	1
		1	8
		2	6

Part 4—Carrying in Column Addition

Add:

1.	7.	8.	11.	12.
29	27	9132	653	
47	96	5112	5156	
	77	2343	7360	
	84	7417	2661	
	98		4099	
	17		9730	
	26		9222	

12. Copy and add: $62 + 604 + 827 + 797 + 997 = ?$ (Put your work under 12 above.)

Part 5—Checking Answers in Addition

Directions: Some of the printed answers below are right, and some are wrong. Check each sum by adding downwards. Write your check answer on the line at the top of the example. The first one is already done correctly.

797		
213	165	6566
128	923	3157
456	928	4918
787	466	7194
	2484	77
		21792

Let us suppose the test has been given and scored and the scores filled in at the top of the page as illustrated:

TEST I: ADDITION OF WHOLE NUMBERS: FORM A

Name *Mary N. Morgan* Grade *4A* Boy or girl? *Girl* Age *9*
When is your next birthday? *Nov. 1* How old will you be then? *10*
School *X* *Bloomington* *Indiana* Date *Oct. 11, 1927*
(Name) (City) (State)

Summary of Pupil's Score	*Part 1*	*Part 2*	*Part 3*	*Part 4*	*Part 5*	*Total*
Number of Examples Right.....	54	50	35	60	22	221
Educational Age Equivalent.....						9–0
Grade Equivalent of Score.....	H3	H3	L4	H3	L4	H3+

H stands for "high"; L stands for "low."

Although Mary is nearly ten years of age, her total score is equal only to that of the average nine-year-old. Although she is in the 4A grade or nearly at the middle of the 4th year's work, comparison with the standards for this test shows that she averages little better in the various phases of the work than the average 3B pupil at the end of the term. (See Grade Equivalent of Score.) This fact suggests the desirability of trying to find out what the child's particular difficulty is. This discovery can be made by going through the test and noting the types of examples missed.

Mary missed in Part 1, Basic Addition Facts, the following combinations: $7+6=$; $8+5=$; $\begin{array}{r} 7 \\ 5 \\ \hline \end{array}$ $\begin{array}{r} 7 \\ 4 \\ \hline \end{array}$ $\begin{array}{r} 9 \\ 7 \\ \hline \end{array}$ $\begin{array}{r} 9. \\ 5 \\ \hline \end{array}$ She did not attempt the last six of the column type and the two-line type at the bottom of the test. Mary's difficulty seems to be primarily with 7, 8, and 9 combinations with numbers that total more than 10. The probable cause of her weakness as revealed by the analysis of this part of the test is that she has never learned these particular combinations and that 7, especially, is a difficult number in combination for her.

Further analysis of her difficulty through the use of informal

tests revealed that with several combinations she was using the type of crutch not uncommon with children who have learned to add fives and tens and know the combinations that make these numbers. She often reasoned as she did in the case of 9 and 9 where she proceeded as follows: 10 and 10 are 20; 9 is one less than 10; 2 from 20 is 18. Here is diagnosis that reveals particular lack of knowledge or incorrect thinking.

The next step in removing difficulty is reteaching. It had been presumed that she knew her number combinations. It was revealed that she did not. The only thing to be done was to teach these combinations and then to drill on them for the purpose of developing automatic responses to these stimuli. This effort, of course, had to be followed by re-testing to determine whether the work had been successful. The summary of the whole process is test, diagnose the results of testing to determine the difficulties encountered, then find the cause of this difficulty; in the light of these reteach and drill upon the processes; and re-test to determine the efficiency of the work.

It is not possible here to go into every phase of diagnosis and remedial work that may be developed by the use of diagnostic tests for individual pupils. However, excellent examples of how diagnosis may reveal the cause of slow work and special types of difficulty are found in the studies of Uhl, Buswell and John, Stevenson, and others.

Analysis of the difficulties encountered by a class as a whole may be made with profit. Let us take the results with the class that took the Compass Arithmetic Test I; let us analyze the results of the class on Part 2 of this test by placing across the top of a large sheet of paper the numbers of the problems and along the left-hand margin the names of the pupils. (See Table 10.) Then, by examining each pupil's paper and indicating with a cross in the appropriate space each pupil's mistakes and by summarizing these results, one is able to get a picture of the condition of the class and to find particular difficulties that may be outstanding for the class as a whole.

There are sixty-six examples in Part 2. Only the first ten and the last six, however, are shown on the chart.

TABLE 10.—PROBLEMS INCORRECTLY WORKED IN PART 2 OF TEST I OF COMPASS ARITHMETIC TEST BY 2B CLASS IN XYZ SCHOOL, INDIANA, 1927

	2 20	1 21	6 22	8 13	18 6	4 12	9 29	7 25	24 7	10 2	20 +7	26 +6	28 +7	18 +7	0 +20	22 +0
John A.								x								
Philip B.							x									
Helen B.								x							x	
Fred D.			x					x					x			
Hazel D.								x							x	
Wilma F.				x			x						x		x	
Eugene G.								x								
LeRoy J.							x									
Mary L.								x	x				x	x	x	
Florence M.			x				x									
John McI.																
Phoebe M.								x								
Jean R.					x			x								
Henry R.							x	x								
Helen W.							x	x								

X indicates that the problem was solved incorrectly.

Although there is a scattering of individual difficulties, at least two or three types of problem offered special difficulties to the class as a whole. 7 was missed by 10 of the 15 members

of the class. Ordinarily this is not a particularly difficult combination, but to this particular class it seems to have offered special difficulty. This may have been due to two reasons: the combination may not have been well known, or the position of the problem with the larger number at the bottom may have caused some difficulty.

Whatever the cause of this deficiency, this combination must be re-taught. To be sure that it is perfectly learned, it will be necessary not only to teach the 25 and 7, but 5 and 7, 15 and 7, 35 and 7, etc., as well as to teach that the position of larger and smaller numbers in the column does not affect the total. This method of diagnosis and remedial work is, of course, suggested on the assumption that children add up and not down and that they add in the fashion 25 and 7 and not 5 and 7 carrying 1 to add to the 2. Sufficient drill must then be given to make sure that these skills or abilities have become habitual.

Selected illustrations from parts of two other tests of the Compass Diagnostic Tests are given:

TEST IV: DIVISION OF WHOLE NUMBERS: FORM A

Part 1—The Vocabulary of Division

Directions: Four possible answers follow each of the first seven exercises below. Draw a line under the *one* of the four possible answers which makes the sentence true.

1. An example of division is: $6+3=9$ $18-4=14$ $8\div 2=4$ $14\times 4=56$
2. In the example, $8\div 4=2$, 2 is the: dividend quotient remainder divisor
3. The sign used to indicate division is: $+$ $\div$ $-$ $\times$
4. The example, $9\overline{)810}$, would be solved by: multiplication division subtraction addition

Part 2—Fundamentals of Short Division

How many 9's in 72?.....	$45\div 5=$	$9\overline{)81}$	$7\div 7=$
How many 8's in 32?.....	$28\div 4=$	$6\overline{)42}$	$0\div 8=$

Part 3—Short Division with Carrying

Directions: Work the next twelve examples. Write only the answers.

$2\overline{)2648}$ $5\overline{)1055}$ $3\overline{)69690}$ $4\overline{)88088}$

Part 4—Multiplication, Addition, and Subtraction Used in Division in Parts 2, 3, 5, 6, and 7

Multiply (Do not divide):

$8\times6=$	$7\times5=$	$3\times8=$	$5\times8=$	$3\times7=$	$2\times7=$
$4\times9=$	$6\times0=$	$7\times0=$	$5\times4=$	$0\times6=$	$3\times6=$
$7\times8=$	$9\times6=$	$5\times2=$	$0\times5=$	$4\times6=$	$1\times5=$

Add (Do not multiply):

$56+6=$	$36+7=$	$8+8=$	$36+2=$	$4+8=$
$21+1=$	$0+1=$	$24+2=$	$5+7=$	$3+5=$

Subtract (Do not add):

$44-42=$	$41-36=$	$35-35=$	$22-21=$
$10-10=$	$10-8=$	$14-12=$	$8-5=$

2550	1518	412	1207	427	4389
-2448	-1480	-412	-1064	-384	-4389
1064	222	3337	6675	282	2513
-1064	-222	-2670	-6230	-240	-2436
205	599	1228	89	32	1020
-156	-560	-1184	-74	-27	-816

Part 5—Estimating the First Quotient Figure

Directions: Write only the first correct quotient figure in its proper place. Do not take time to finish the examples. Look at the samples before you begin to work.

SAMPLES: $8\overline{)896}$ (1) $87\overline{)563}$ (6)

$48\overline{)282}$ $74\overline{)222}$ $9\overline{)32}$ $296\overline{)447}$ $64\overline{)195}$ $408\overline{)1020}$ $9\overline{)2}$

$560\overline{)3920}$ $89\overline{)667}$ $627\overline{)796}$ $40\overline{)7}$ $89\overline{)333}$ $627\overline{)1692}$ $48\overline{)427}$

Part 6—Fundamentals of Long Division; Checking

Directions: Work the eleven examples below.

1.	2.	3.	4.
9)21242	64)1955	74)22289	40)87207

Part 7—Finding Errors in Long Division

Directions: Study the five examples below. Some of the examples have mistakes in them. You are to put a heavy line through the first mistake you find in each example, as shown in the sample.

SAMPLE:

```
      20
213)4260
   426
   ---
   100
```

1.

```
   454
266)120764
   1064
   ----
   1436
   1330
   ----
   1064
   1044
   ----
     20
```

2.

```
   276
78)20528
   156
   ---
   592
   546
   ---
   468
   468
   ---
```

TEST XV: INTEREST AND BUSINESS FORMS: FORM A

Part 1—Vocabulary of Interest and Business Forms

Directions: Underline the best answer among the several choices given to complete each statement.

1. Interest is reckoned by the day week month year quarter
2. A common rate of interest on savings accounts is 1/10 of 1% 4% 7% 15% 85%
3. A reduction on a bill for cash payment is called a commission discount check stub balance
4. In order to cash a check it must be endorsed discounted receipted balanced sold
5. A stub is attached to a discount receipt check commission deposit slip
6. Par value means face value market value highest price lowest price average price
7. A deposit is money taken out money put in the balance a commission a good security

A computation test for individual pupils which allows for a more detailed analysis of individual difficulties is "Diagnostic Chart for Individual Difficulties, Fundamental Processes in Arithmetic." [10] This is, as a matter of fact, not a test but a chart for diagnosing difficulties for use with individual pupils who are doing unsatisfactory work in arithmetic. The chart covers the four fundamental processes of addition, subtraction, multiplication, and division. Only one process is diagnosed at a time. The pupil is seated at a table where the work is to be done and is provided with a work sheet as shown below. He is told to work the examples in the way he ordinarily does. He is also told to do as much of his work as he can out loud, so the teacher can understand how he gets his answers. The teacher does not make suggestions, nor aid the pupil in any way. The teacher desires to determine just how the pupil is doing his thinking and, if there are difficulties encountered as indicated by incorrect answers or procedures, the teacher wishes to know the cause of these difficulties. The teacher uses a diagnosis sheet as shown on page 99 and on it indicates the pupil's method of thinking and the types of mistakes made. The mistakes may be of two general types—those that indicate incorrect methods although the answer is correct, and another type where incorrect thinking produces incorrect answers. Typical examples are given on pages 99 and 100.

PUPIL'S WORK SHEET [11]

School..... Name.....

Add:

(1)	5	6	(2)	2	8	(3)	12	13
	2	3		9	4		2	5
(4)	19	17	(5)	6+2=		(6)	52	40
	2	9					13	39
				3+4=				

[10] By G. T. Buswell and Lenore John, published by Public School Publishing Company.

[11] From the Diagnostic Chart for Fundamental Processes in Arithmetic, by G. T. Buswell and Lenore John.

Add:

(7) 78 71 — 46 92 —

(8) 3 5 8 2 — 8 7 9 7 —

(9) 53 8 — 7 89 —

Subtract:

(1) 5 3 — 8 8 —

(2) 7 − 1 =
9 − 0 =

(3) 19 2 — 15 4 —

(4) 58 4 — 79 3 —

(5) 36 21 — 79 24 —

(6) 12 6 — 10 2 —

(7) 15 13 — 19 12 —

(8) 59 − 2 =
86 − 4 =

(9) 346 215 — 836 302 —

Multiply:

(1) 4 2 — 3 3 —

(4) 343 2 — 2212 4 —

(7) 52 3 — 91 7 —

(10) 54 5 — 43 8 —

(13) 35897 2 — 76097 8 —

(16) 52 30 — 769 80 —

(19) 53 321 —

(21) 512 231 —

Divide:

(3) $3\overline{)693}$ $2\overline{)46}$

(6) $3\overline{)1209}$ $8\overline{)5608}$

(9) $23\overline{)483}$ $32\overline{)992}$

(12) $2\overline{)6140}$ $6\overline{)54036}$

(15) $21\overline{)1491}$ $34\overline{)16354}$

(17) $5\overline{)151}$ $6\overline{)204183}$

(19) $37\overline{)15133}$ $46\overline{)231104}$

(21) $924\overline{)580272}$ $395\overline{)302175}$

DIAGNOSTIC CHART FOR INDIVIDUAL DIFFICULTIES[12]

Addition (Place a check before each habit observed in the pupil's work):

.... a1 Errors in combinations
.✓. a2 Counting
.... a3 Added carried number last
.... a4 Forgot to add carried number
.... a5 Repeated work after partly done
.... a6 Added carried number irregularly
.... a7 Wrote number to be carried
.... a8 Irregular procedure in column
.✓. a9 Carried wrong number
....a10 Grouped two or more numbers
....a11 Splits numbers into parts
....a12 Used wrong fundamental operation
....a13 Lost place in column
....a14 Depended on visualization
....a15 Disregarded column position
....a16 Omitted one or more digits
....a17 Errors in reading numbers
....a18 Dropped back one or more tens
....a19 Derived unknown combination from familiar one
....a20 Disregarded one column
....a21 Error in writing answer
....a22 Skipped one or more decades
....a23 Carrying when there was nothing to carry
....a24 Used scratch paper
....a25 Added in pairs, giving last sum as answer
....a26 Added same digit in two columns
....a27 Wrote carried number in answer
....a28 Added same number twice

Habits not listed above.....

a2. Counting[13]

Example 1:

$$\begin{array}{r} 8 \\ 7 \\ 9 \\ 7 \\ \hline 31 \end{array}$$

This pupil counted by ones to get any combination he did not know. He tapped with his pencil as he counted, usually in

[12] *Ibid.*
[13] *Ibid.*

groups of threes. In working the example above he did as follows:

15 and 9 are—(tapping) 16 19 22
17 20 23
18 21 24

24 and 7 are—(tapping) 25 28 31
26 29
27 30

When asked to make his taps in groups of fours and to proceed horizontally instead of vertically, he did so with little effort. He was also able to get the answer by watching some one else tap, even when the taps were ungrouped.

a9.
17
9
—
71

7 and 9 is 16, 6 and 1 is 7
Carried wrong number

The types of mistakes are then appropriately indicated at the top of the page before the "Habits Observed in the Pupil's Work." Such diagnosis as this is basic to overcoming the particular difficulty of individual pupils. Its use by a skilled teacher can contribute much to the effective learning by children below the average in achievement in arithmetic. Such a diagnostic chart, concerned primarily with the written or reasoning problems in arithmetic, would be a distinct contribution to the field of measurement and diagnosis. It must, of course, be recognized that the possibilities of some children in arithmetic are limited by the lack of intelligence; however, probably most of the difficulty in arithmetic is due to poor learning and can be remedied.

Some studies have been made of types of errors children make in the solution of written arithmetic problems, and types

of teaching and drill to overcome these particular difficulties. Morton [14] lists the following types of errors observed:

1. Procedure wholly wrong or entirely inadequate.
2. Procedure partly wrong.
3. Inadequate knowledge of the processes involved.
4. Procedure correct so far as carried out, but incomplete.
5. Errors in computation.
6. Error in copying figures or in transcribing words into figures.
7. Lack of knowledge of facts.
8. Unknown.

The percentages of the types of errors are given in Table 11.

TABLE 11.—THE ERRORS CLASSIFIED AND FREQUENCIES CONVERTED INTO PER CENTS

Grade	*Classes of Errors*								*Total*
	1	2	3	4	5	6	7	8	
V.....	69.5	9.4	3.8	4.8	7.4	4.4	0.0	0.6	99.9
VI.....	57.2	5.2	8.9	3.7	16.3	5.5	0.9	2.2	99.9
VII.....	47.5	10.0	4.6	9.2	21.7	5.4	0.8	0.8	100.0
VIII.....	33.3	12.1	8.3	15.9	27.3	2.3	0.8	0.0	100.0
TOTAL..	57.8	8.7	5.9	6.6	14.9	4.7	0.5	1.0	100.1

This table may be read as follows: In Grade 5, 69.5 per cent of the errors were of type 1, listed "Procedure wholly wrong or entirely inadequate"; 9.4 per cent of type 2, "Procedure partly wrong"; 3.8 per cent of type 3, "Inadequate knowledge of the processes involved," etc.

This study of errors in problems seems to point to a very distinct need of a diagnostic type of test measuring the specific difficulties of the pupil who does not seem to understand the procedure of solving arithmetic problems. This particular

[14] *An Analysis of Errors in the Solution of Arithmetic Problems,* Educational Research Bulletin, Vol. IV, No. 9 (April 29, 1925), pp. 187-90. Also found in Morton's *Teaching Arithmetic in the Intermediate Grades,* Silver, Burdett and Company.

cause alone accounts for more than one-half of the errors. Remedial teaching and drill work in this line are not yet as well developed as they should be. Stevenson [15] goes somewhat further into the particular difficulties of children and notes the following important causes of failure: (1) physical defects; (2) lack of mentality; (3) lack of skill in fundamentals; (4) inability to read, which, of necessity, affects the ability to read arithmetic problems; (5) lack of general and technical vocabulary; and (6) lack of proper methods or technique for attacking problems. He suggests the following remedial measures:

I. Teach pupils to read problems and develop a technique for working them. When assigning the lesson, have the pupils read the problem silently and teach them to pick out (1) what they are asked to find, (2) what is given to help answer the question, (3) what process or processes are to be used, and instruct them to estimate answers.

II. Teach the vocabulary used in problems. The pupils should state the problems in their own words, and in as many different ways as possible, while special attention should be given to technical words such as rate, salary, per cent, dealer, commission, etc.

III. Dramatize problems referring to measurements, e.g., pints, quarts, inches, feet, areas, etc.

IV. Give a large variety of problems from life situations. The pupils should be encouraged to make up problems dealing with their own or their parents' activities. An excellent device to increase the ease of understanding problems is to vary the language in which they are stated. Problems centering around specific activities such as a field meet, a picnic, etc., arouse the interest of the pupils and serve their arithmetical purpose, too.

V. Give individual instruction. An excellent rule in arithmetic problems, as in other subjects, is to find out what the pupil needs and see that he gets it. For example, if he has vocabulary difficulties and cannot estimate answers, give him special instruction in these two phases.

[15] "Difficulties in Problem Solving," *Journal of Educational Research*, Vol. XI, No. 2, February, 1925, pp. 95-103.

Several other writers have shown the value of remedial drill of the types suggested.

In this chapter we have attempted (1) to state the objectives of the teaching of arithmetic as they are found in the more up-to-date courses of study, (2) to analyze some of the abilities and skills involved in arithmetic computation and in problem solving, (3) to describe some survey and diagnostic tests, (4) to point out the need for diagnostic work in problem solving, (5) to show how the results of arithmetic tests may be used as a basis for remedial work, (6) to suggest types of difficulties ordinarily revealed by diagnostic testing, and (7) to suggest remedial procedures.

Following is a list of tests which have been found satisfactory for use in measuring achievement in arithmetic.

TESTS IN ARITHMETIC

Buswell-John Diagnostic Test for Fundamental Processes in Arithmetic. By G. T. Buswell and Lenore John. Published: 1926. Purpose: to test and to diagnose individual difficulties in arithmetic. Range: elementary school. Time to give: 15 or 20 minutes. Time to score: short. Standardized: as to procedure and use.
Publisher: Public School Publishing Co.
Price: $2.50 per package of 50 tests; specimen set, 20 cents.

Compass Diagnostic Tests in Arithmetic. By G. M. Ruch, F. B. Knight, H. A. Greene, and J. W. Studebaker. Published: 1925. Purpose: to test the basic arithmetic processes and skills commonly taught in grades 2-8 and to provide diagnosis for individual pupils. Range: grades 2-8. Time to give: one class period for each test. Time to score: short. Standardized.
Publisher: Scott, Foresman and Co.
Price: from 25 cents to $1.25 per package of 25 tests according to the tests chosen; specimen set, 60 cents.

Compass Survey Tests in Arithmetic. By H. A. Greene, F. B. Knight, G. M. Ruch, and J. W. Studebaker. Published: 1927. Purpose: to measure achievement in arithmetic. Range: grades 2-8. Time to give: elementary examination, 25 minutes; advanced examination, 35 minutes. Time to score: short. Standardized.
Publisher: Scott, Foresman and Co.

Price: 50 cents per package of 25 tests, either form; Manual of Directions, 15 cents.

Courtis Standard Supervisory Tests in Arithmetic. By S. A. Courtis. Published: 1918. Purpose: to measure pupils' ability in the fundamentals of arithmetic. Range: Test A, grades 4B, 4A, and 5B; Test B, grades 5A-8A. Time to give: Test A, 3 to 6¼ minutes. Test B, 4½ to 9 minutes. Time to score: no information. Standardized.
Publisher: S. A. Courtis, 1807 East Grand Boulevard, Detroit, Michigan.
Price: specimen set, 32 cents, complete.

Foran Diagnostic Computation Scale. By T. G. Foran. Published: 1925. Purpose: to measure a pupil's ability in the fundamentals of arithmetic and diagnose and classify errors. Range: Part I, grades 2-4; Part II, grades 5-8. Time to give, 25 minutes for Part I; 30 minutes for Part II. Time to score: about 2 minutes. Standardized.
Publisher: The Catholic Education Press.
Price: $1.55 per package of 100 tests, including three manuals of directions.

Los Angeles Diagnostic Tests: Fundamentals of Arithmetic. By Caroline Armstrong and W. W. Clark. Published: 1925. Purpose: to provide a standardized survey test in the four fundamentals and to determine individual errors so that remedial treatment may be given. Range: grades 2-8. Time to give: 40 minutes. Time to score: no information. Standardized.
Publisher: Research Service Co.
Price: $3 per 100; specimen set, 10 cents.

Los Angeles Diagnostic Tests: Reasoning in Arithmetic. By Caroline Armstrong. Published: 1926. Purpose: to measure achievement and diagnose difficulties in arithmetic. Range: grades 3-9. Time to give: grades 3-5, 30 minutes; grades 6-9, 40 minutes. Time to score: 1 minute. Standardized.
Publisher: Research Service Co.
Price: $3 per 100 copies, any form; $25 per 1,000; specimen set, complete, 10 cents.

Monroe's General Survey Scale in Arithmetic. By Walter S. Monroe. Published: 1920. Purpose: to measure pupils' ability to perform the operations of arithmetic. Range: Scale I, grades 3-5; Scale II, grades 6-8. Time to give: Scale I, 7 minutes; Scale II, 17½ minutes. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: $1 per 100; specimen sets, 15 cents.

Monroe's Standardized Reasoning Tests in Arithmetic. By W. S. Monroe. Published: 1918. Purpose: to measure ability to reason in arithmetic. Range: grades 4-8. Time to give: 25 minutes. Time to score: about 1 minute. Standardized.
Publisher: Public School Publishing Co.
Price: 80 cents per 100, each part; specimen set, 8 cents.

The National Arithmetic Scales. By J. Freeman Guy. Published: 1927. Purpose: to measure achievement in the 4 fundamentals. Range: grades 3-9. Time to give: no information. Time to score: short. Standardized.
Publisher: National Publishing Society.
Price: 1 cent per test sheet; directions, 6 cents; class record sheets, 2 cents each.

Otis Reasoning Test in Arithmetic. By Arthur S. Otis. Published: 1918, 1923. Purpose: to test pupils' reasoning ability in arithmetic. Range: grades 4-9. Time to give: 5 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: 50 cents per package of 25 examination sheets of either form, complete; specimen set, 10 cents, postpaid.

Reavis-Breslich Diagnostic Tests in the Fundamental Operations of Arithmetic and in Problem Solving. By W. C. Reavis and E. R. Breslich. Published: 1927. Purpose: to diagnose individual work in arithmetic and to be used as a means of classification. Range: grades 7-9. Time to give: 30 minutes. Time to score: about 5 minutes. Standardized.
Publisher: University of Chicago Press.
Price: 10 cents per copy; 10 cents for directions, etc., with usual discount for schools.

The Spencer Diagnostic Arithmetic Test. By Peter L. Spencer. Published: 1924. Purpose: to measure achievement and diagnose pupil difficulties. Range: grades 3-8. Time to give: 45 minutes. Time to score: 5 minutes. Standardized.
Publisher: Bureau of Educational Research, University of Oregon.
Price: $2 per 100; specimen sets, 10 cents for each form, postpaid; 6 forms, 45 cents.

Stevenson Problem Analysis Test. By P. R. Stevenson. Published: 1923. Purpose: to test the ability of pupils to analyze problems in arithmetic before making numerical computations. Range: Test I, grades 4-6; Test II, grades 7-9. Time to give: no time limit. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: $1 per 100; specimen set, 10 cents.

New Stone Reasoning Test in Arithmetic. By C. W. Stone. Published: 1927. Purpose: to measure achievement and diagnose difficulties in the reasoning of arithmetic. Range: grades 4-9. Time to give: 30 to 40 minutes. Time to score: 4 minutes. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: $1.25 per 100 copies; $10 per 1,000; Manual of Directions, 65 cents; specimen set, complete, 75 cents.

Wilson Inventory and Diagnostic Tests in Arithmetic. By Guy M. Wilson. Published: 1927. Purpose: to serve as complete inventory of essential facts and process difficulties in four fundamental processes and fractions. Range: grades 3-10. Time to give: no time limit. Time to score: 1 to 10 minutes for each test. Standardized.
Publisher: University Publishing Co.
Price: not yet determined.

Wisconsin Inventory Tests in Arithmetic. By Worth J. Osburn. Published: 1924. Purpose: to test pupils' fund of arithmetical knowledge. Range: grades 2-8. Time to give: no time limit. Time to score: 1 minute. Standardized: not a survey test, therefore standards are not emphasized.
Publisher: Public School Publishing Co.
Price: $1 per 100 for each test; specimen set, all 8 tests, 10 cents.

Woody-McCall Mixed Fundamentals. By Clifford Woody and W. A. McCall. Published: 1916. Purpose: to measure achievement in the fundamentals. Range: grades 3-8. Time to give: 20 minutes. Time to score: 1 minute. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: 60 cents per 100 copies; $5.50 per 1,000; specimen set, 20 cents, postpaid.

BIBLIOGRAPHY

Anderson, C. J. "Use of the Woody Scale for Diagnostic Purposes," *Elementary School Journal,* Vol. 18: pp. 770-81, June, 1918.

Ballenger, H. L. "Overcoming Some Addition Difficulties," *Journal of Educational Research,* Vol. 13: pp. 111-17, February, 1926.

Brueckner, L. J. "Certain Arithmetic Abilities of Second Grade Pupils," *Elementary School Journal,* Vol. 27: pp. 433-43, September to June, 1926-27.

Buswell, Guy Thomas and John, Lenore. "Diagnostic Studies in Arithmetic," *Supplementary Educational Monograph,* No. 30, University of Chicago, July, 1926.

Buswell, Guy Thomas. "Summary of Educational Investigation Relating to Arithmetic," *Supplementary Educational Monograph,* No. 27, Department of Education, University of Chicago, June, 1925.

Davies, G. R. "Elements of Arithmetical Ability," *Journal of Educational Psychology,* Vol. 5: pp. 131-40, March 5, 1914.

Hunkins, R. V. and Breed, F. S. "The Validity of Arithmetical Reasoning Tests," *Elementary School Journal,* Vol. 23: pp. 453-66, February, 1923.

Kelley, F. J. "The Results of Three Types of Drill on the Fundamentals of Arithmetic," *Journal of Educational Research,* Vol. 2: pp. 693-701, November, 1920.

Monroe, Walter S. "The Derivation of Reasoning Tests in Arithmetic," *School and Society,* Vol. 8: pp. 324-29, September 14, 1918.

Morton, R. L. "Analysis of Pupils' Errors in Fractions," *Journal of Educational Research,* Vol. 9: pp. 117-25, July, 1924.

Morton, R. L. "Pupils' Errors in Solving Arithmetic Problems," *Educational Research Bulletin,* Ohio State University, Vol. 14: No. 8, p. 155, April 15, 1925.

Morton, R. L. "Solving Arithmetic Problems: Case Studies," *Educational Research Bulletin,* Ohio State University, Vol. 14, No. 10: p. 203, May 13, 1925.

O'Brien, F. P. "Improvement of Instruction in Arithmetic," *Bulletin of Education,* Vol. 1, No. 4: University of Kansas, October, 1927.

Rimer, Wilson, and Knight. "How Pupils Solve Problems in Arithmetic," *Educational Research Bulletin,* Ohio State University, Vol. 3, No. 13: p. 274, October 15, 1924.

Stevenson, P. R. "Increasing the Ability of Pupils to Solve Arithmetic Problems," *Educational Research Bulletin,* Ohio State University, Vol. 3, No. 13: p. 267, October 15, 1924.

Thorndike, E. L. "The Relation Between Speed and Accuracy in Addition," *Journal of Educational Psychology,* Vol. 5: pp. 537-541, November, 1914.

Uhl, W. L. "The Use of Standardized Materials in Arithmetic for Diagnosing Pupils' Methods of Work," *Elementary School Journal,* Vol. 18: pp. 215-19, November, 1917.

Upton, C. B. "Influence of Standardized Tests on the Curriculum in Arithmetic," *Teachers College Record,* Vol. 26: pp. 627-41, April, 1925.

Wilson, Estaline. "Improving the Ability to Read Arithmetic Problems," *Elementary School Journal,* Vol. 22: pp. 380-86, January, 1922.

CHAPTER VI

Measurement in Secondary Mathematics

Algebra and geometry form an important part of, and exert an outstanding influence on, the secondary school curriculum. Much of the time and effort of our secondary school pupils is given to learning the skills, principles, and concepts of these subjects. They are closely allied to many phases of the science and the technical work of the secondary school. They are usually required for graduation. These subjects form a rather constant unit of the high school work required for college entrance. Mathematics is playing an increasingly important part in the recent development of scientific research, in industrial and commercial development, and in the work of the professions. Within the last few years these subjects have been the center of controversy among those interested in the development of the secondary school curriculum. One controversy has been between those who favor a blanket requirement of algebra and geometry for all pupils and those who favor the inclusion of these subjects only as electives; another controversy concerns attempts to develop the content of these courses in terms of known social values so as to make them contribute primarily to the aims of vocational and social studies against attempts to make their content contribute as preparatory to advanced work. Still another point of view is that the contribution of these subjects is primarily through their disciplinary values.

Unquestionably the content of these subjects and emphasis upon them are much influenced by college entrance requirements on one side and an increasingly unselected high school

population on the other. The more difficult problems of the teaching of algebra and geometry are raised in connection with the instruction of that group of pupils constituting the lower 40 per cent of ability and are closely related to a study of the laws of mental growth and development of this group. Due to the development of the junior high school and the development of a curriculum to meet its special needs, some of the content of algebra and geometry originally taught in the high school is now being taught in the seventh and eighth grades in the new junior high school courses in mathematics.

It is often characteristic of these subjects that their objectives are given in terms of subject objectives rather than of life objectives. That is, they are most often given in terms of their technical or propædeutic values rather than in terms of their social values. In many cases, of course, these are identical. The following statement of objectives [1] in algebra would be generally accepted.

1. To develop the ability to further understand the meaning and significance of arithmetic.
2. To develop the ability to interpret correctly and to construct graphs.
3. To develop understanding of the "language" of algebra.
4. To develop the ability to understand, solve, transform, and apply formulæ.
5. To develop skills and techniques that are used in the solution of examples and problems and to apply these to the solution of new problems.
6. To develop the ability to see equations in life situations and to solve equations.

What is the desirable content of ninth grade algebra? The answer to this question should of course be determined in the

[1] The functions of mathematics in secondary education are set forth in detail in *The Reorganization of Mathematics in Secondary Education*, a summary of the report by the National Committee on Mathematics Requirements, Bulletin No. 32, U. S. Bureau of Education, Washington, D. C., 1922.

light of the objectives of teaching. If the objectives just stated were accepted, much of the present content of algebra would disappear. Back of this study of objectives and content, and more fundamental to the development of the science of testing, diagnosis, and remedial work in algebra, is the question: "What are the skills, abilities, knowledge of rules, and concepts that are fundamental to achievement in the various units of content taught in algebra?" One example will illustrate:

Ability with equations includes two groups of abilities which are, at least psychologically, very different. The first is to manipulate the equation so as to obtain a numerical value for the literal element, or so as to obtain a value for one of the literal elements in terms of the others. The equation is "solved." The second is to understand the equation as the expression of a certain relation whereby we can correctly prophecy what value a certain element will have, according to the values which one or more other elements have. Thus $\frac{Q}{2} = KR + 4$ is "solved" for Q by finding that $Q = 2KR + 8$. $\frac{Q}{2} = KR + 4$ is "understood with respect to the relation between Q and R" by understanding that if K is a constant, Q is in direct proportion to R, that if Q is expressed as ordinate to fit R abscissa value, we have a set of points on a straight line cutting the y axis at $+8$, with a slope depending on what K is, and that every increment added to R produces, other things being equal, an increment of $2K$ in Q. The older algebra neglected the second ability almost entirely, and even yet the first ability is given far more time and attention in most textbooks, courses, and examinations. Yet the second ability seems of equal or greater importance.[2]

Measurement in the field of algebra has been deficient in two respects: (1) it has attempted to measure only the fundamentals or the mastery of the mechanics of mathematical pro-

[2] E. L. Thorndike (and others), *The Psychology of Algebra*, The Macmillan Co., 1922, p. 105: "Ability with Equations."

cedure and, accordingly, has not distinguished between the social, vocational, and propædeutic aims of the teaching of algebra; (2) it has not attempted the analytical and diagnostic possibilities revealed by study of the psychology of algebra and of the thinking of adolescent boys and girls. The ideal diagnostic or inventory type of measurement that could be developed out of such an analysis of skills and abilities would test each ability not only in isolation but also in the context in which it is most commonly used. Such testing in the light of the understanding of boys and girls of adolescent age would be a real basis for remedial teaching.

Testing in the field of algebra should include: (1) inventory tests, (2) survey achievement tests in algebra computation and problem solving devised to give speed and accuracy measures if desired, (3) diagnostic computation test, (4) diagnostic problem solving test, (5) individual diagnosis charts in computation, (6) individual diagnosis chart in problem solving, (7) practice test or drill pads, (8) test of social traits and disciplinary values in algebra.

ALGEBRA TESTS

A power test by Hotz [3] consists of five scales: (1) Addition and Subtraction, (2) Multiplication and Division, (3) Equation and formula, (4) Problems, (5) Graphs. This test measures to some extent objectives 2, 4, 5, and 6 stated on page 110.

The examples and problems of the test are of increasing difficulty and, on Series *A*, the difficulty increases by equally spaced steps. These steps have been determined statistically. The first problem in each scale is very easy and the last one very difficult. The items included, according to the author,

[3] *First Year Algebra Scales*, Teachers College, Columbia University, Contributions to Education No. 90, 1918. There are two series of these tests, *A* and *B*. Series *B* includes all the tests, Series *A*, a selected and graduated group of tests.

are those most commonly found in recent textbooks.[4] The norms are derived from a large number of cases and are available for different times of the school year.

The tests may be used as follows:

After 3 months' teaching	Addition and Subtraction Equation and Formula
After 6 months' teaching	Multiplication and Division Problem
End of 1 year's teaching	Equation and Formula Graph

A few extracts from the tests[5] will show the nature of the problems and examples used.

Addition and Subtraction

Carefully perform the operations as indicated.

(1) $4r+3r+2r=$ (5) $7x-x+6-4=$

(22) $\dfrac{3-2x}{(x-1)^3}+\dfrac{x+1}{(x-1)^2}-\dfrac{1}{(x-1)}=$

Problems

(2) A man is m years old; how old was he r years ago?

Answer.....

(8) A rectangular box is d inches deep, w inches wide, and contains r cubic inches. What is its length?

Answer.....

Instructions for giving and scoring the tests are clear and should be followed exactly. The reliability of the scales is .92 for Series *A*. The validity of the test as measured by teachers' marks is generally not better than .50, and as measured with other algebra tests is usually not higher than .70.

[4] *Ibid.*

[5] Published by the Bureau of Publications, Teachers College, Columbia University.

It is ordinarily assumed in the use of this test that the achievement represented by it is the result of the teaching of the class in algebra. Some of the test could, however, be done by pupils who have finished the seventh and eighth grades. For example, in the part of the test on graphs, Series *A,* probably the first four of the eleven items can be answered by most eighth grade pupils even in an eight-four organization and certainly could be answered by most eighth grade pupils in a six-three-three organization. The validity of the results from the Hotz test as a measure of the results of teaching would have to be interpreted in terms of an inventory test. The content of the test is based upon the analysis of textbooks published before 1918. The textbooks of more recent years have changed some in their content and especially tend to emphasize the more definitely social values of algebra.

The achievement of pupils in algebra is, of course, very much affected by their intelligence. This is proved by the rather high correlation ordinarily found between the results of a measure of ability in algebra and a measure of intelligence. Therefore the norms of algebra tests need to be developed in terms of ability. Hotz's test needs to be re-worked or re-evaluated in terms of modern methods of teaching. It is to be said in favor of the Hotz Scale that it is relatively easy to use and tests a wide range of abilities rather accurately. Its use as a survey test is recommended. Only one form of the test is available. Forms A and B are not equated forms.

An example of the partially diagnostic type of test is the Douglass Standard Diagnostic Tests for Elementary Algebra.[6] The first step in the development of this group of tests was an attempt to determine what constitutes the fundamental operations in algebra. To determine this, the author asked members of the Mathematical Association of America to designate these fundamental operations in first-year algebra. The test consists of Series *A,* Test I, Addition and Subtraction, Test II, Multi-

[6] Published by the University of Oregon.

plication, Test III, Division, Test IV, Solution of Simple Equations; Series *B*, Test I, Fractions, Test II, Factoring, Test III, Fractional Equations, Test IV, Simultaneous Equations, Test V, Graphs, Test VI, Square Root, Radicals, Exponents, Test VII, Quadratic Equations. This test measures to some degree objectives 2, 4, 5, and 6 stated on page 110.

The content of the tests was based upon these four principles:[7]

1. The exercises selected should clearly require proficiency in the fundamental process for which the test was being constructed.
2. The list of exercises in each test should provide for testing the chief subtypes of difficulty and teaching units in each fundamental process.
3. The exercises should be so selected that a differentiation of power would be possible on the basis of the degree of difficulty involved.
4. For the purpose of complete measurement and differentiation each test should contain one or more exercises which could be solved by only a small per cent of first-year algebra pupils.

An example of the test is found in the following extracts from tests of both Series *A* and *B*:

SERIES A—FORM I

Test I—Addition and Subtraction

1. Add:
$8b$
$3b$
$=$ Ans.

2. Add:
$6x$
$-2x$
$=$ Ans.

3. Add:
$4b^2d$
$-6b^2d$
b^2d
$-4b^2d$
$=$ Ans.

[7] Harl R. Douglass, "A Series of Standardized Diagnostic Tests for the Fundamentals of Elementary Algebra," *Journal of Educational Research*, Vol. IV, No. 5, Dec., 1921, p. 399.

4. Find the sum of: $5c - 4d^2$, $3d^2 - 4c - 3m^2$, and $2d^2 - c - xy$.

5. Simplify by collecting terms: $13a^2 + 7ab^3 - 4x^2y - 10a^2 - 4ab^3 - 2a^2 - 2x^2y + m^3$.

Test II—Multiplication

1. Multiply:
$$\begin{array}{r} 3b \\ 2 \\ \hline \end{array}$$
$= \text{Ans.}$

2. $2m \cdot 5m = ?$

3. Multiply: $2x$ by $-3x$

4. Multiply:
$$\begin{array}{r} 3a^3x \\ 2a^2x^2 \\ \hline \end{array}$$

5. Multiply:
$$\begin{array}{r} 4b^{2x} \\ 3b^x \\ \hline \end{array}$$

6. $6mn\,(7 - 2ma^2) = ?$

SERIES B—FORM I

Test III—Formulæ and Fractional Equations

1. Using the formula $k = \frac{1}{2}bh$, solve for k when $b = 4$ and $h = 3$.
2. Express the following fact in a formula: The interest due on a loan may be found by multiplying the principal by the rate of interest and the number of years the money is loaned.
3. Using the formula below, what does a equal in terms of b and c?

$$c = \frac{a + b}{2}$$

Test V—Graphs

(Note—Students should work exercises 2, 3, and 4 on graph paper, one sheet of which should be at hand before pupil begins tests.

2. Represent by bar diagram the average annual rainfall for the following cities as given opposite each.

San Diego, Cal.	10 inches
Denver, Colo.	14 inches
Springfield, Ill.	37 inches
Dubuque, Ia.	34 inches
Baltimore, Md.	43 inches

An idea of the value of the use of these scales for diagnostic purposes may be obtained from the summary of use made of them by Fossler,[8] who used Series *A* to find the errors in addition, subtraction, and multiplication made by pupils who had completed first-year algebra. An attempt was made in this study not only to find the problems missed but also to find the types of errors made and their frequency. The following two tables give examples of the type of results obtained.

TABLE 12.—PER CENT OF VARIOUS TYPES OF ERRORS MADE BY PUPILS WHO HAD FINISHED FIRST-YEAR ALGEBRA. ADAPTED FROM FOSSLER

Errors in	*Per Cent of Total Errors*
Exponents	26.9
Signs	24.4
Operation of problems	23.1
Coefficients	9.7
Terms	5.9
Miscellaneous	5.7
Letters	4.3

These results are valuable as a picture of conditions revealed by the test. An analysis of such results as these may point to specific needs not only of particular pupils but of the class as a whole. Such an analysis also suggests to teachers of algebra the pitfalls confronting pupils in the study of algebra computation problems. These pitfalls of course should be kept in mind in the teaching of algebra. [See Table 13, page 118.]

[8] "A Study of the Errors Made by Students Who Have Completed First Year Algebra as Shown by the Douglass Standard Diagnostic Tests for Elementary Algebra," Unpublished Master's Thesis, University of Iowa, 1924.

TABLE 13.—NUMBER AND KIND OF ERRORS MADE IN EXPONENTS BY PUPILS FINISHING FIRST-YEAR ALGEBRA AS REVEALED BY DOUGLASS DIAGNOSTIC TEST IN ALGEBRA, SERIES A

	Addition and Subtraction											*Multiplication*											*Division*											*Total in all three tests*
	Problem No.										*Total*	*Problem No.*										*Total*	*Problem No.*										*Total*	
	1	2	3	4	5	6	7	8	9	10		1	2	3	4	5	6	7	8	9	10		1	2	3	4	5	6	7	8	9	10		
Used wrong process		1	5	1		2			1		10				2	3			25	3	78	111				2		2	5			1	10	131
Error in process															1			1	10	21	42	75				1			12		1		14	89
Did not perform process													6	5	1	1		3	47	1	7	71					2		21	3	2	3	31	102
Omitted, Wrong or Disregarded			1	4							5					5	2	8	5	7	27						4	15				19	51
Misread																10					10							9				9	19
Combined unlike letters																	1			89	90													90
Miscellaneous																14				86	100													100
Copied incorrectly			1	1	3	8			1		14						2	4	1		7								2		1	3	24
TOTAL.....		1	7	6	3	10			2		29		6	5	4	33	1	8	94	31	309	491			3	2	6	62	5	3	5	86	606

According to Table 13, one pupil used the wrong process in attempting to solve the second problem in the test in addition and subtraction, and five of the seven pupils who missed the third problem in this test did so for the same reason. In like manner an analysis is shown of the types of mistakes made by other pupils who took the Douglass Diagnostic Test in Algebra.

It should be remembered in the use of the Douglass test that not all abilities and skills in algebra are measured by it. Although a fair sampling of abilities and skills may be made by its use, certainly many are omitted. A truly diagnostic test will have to be more extensive than this one. The important thing for the remedial work following diagnostic testing is knowing the reason for a certain mistake. This takes the teacher back to the pupil's thinking and to the psychology of the subject. Let us take this simple example:

Subtract:

$$\begin{array}{r} -4m^2n \\ 2m^2n \\ \hline = \text{Ans.} \end{array}$$

Many pupils obtain the answer $-2m^2n$. This may result from arithmetic subtraction or algebraic addition. The teacher must know the method that the child used to get the answer before she can straighten out his difficulty. A study of diagnostic testing in algebra shows a decided need for the development of a diagnostic chart or test like the one of Buswell and John in arithmetic, discussed in Chapter V. It is only by means of such a chart in algebra that one can get back to the difficulties or mistakes in thinking that can be overcome by remedial teaching. Osburn has made some analysis of the cause of failure in mathematics. His findings, which follow, are the result of a practical analysis and are worthy of consideration. According to Osburn [9] the following ten phases of the teaching of algebra need to be supplemented to prevent failure:

[9] "Ten Reasons Why Pupils Fail in Mathematics," *The Mathematics Teacher*, Vol. XVIII, No. 4 (April, 1925), pp. 234-38.

1. Exercises in vocabulary.
2. Exercises to develop the meaning of symbols.
3. Training in silent reading to eliminate failure to see and use all of the data.
4. Training in the use of inverse relations.
5. Training to prevent harmful transfer.
6. Drill in horizontal addition and subtraction.
7. Training in reading between the lines.
8. Help for the pupil when conditions seem contradictory.
9. Training pupils to generalize.
10. Proportion.

GEOMETRY TESTS

In the field of geometry the specific aim of the teaching is very much confused by the tendency to mix intuitive and demonstrative geometry. Much of intuitive geometry is now found in the seventh, eighth, and ninth grades of the reorganized mathematics of the junior high school.

In an analysis of topics of thirteen series of textbooks for grades seven, eight, and nine the following percentages of pages were given to intuitive and constructive geometry: [10]

Seventh grade, 12.24 per cent. This per cent is as high as the per cent for any other one topic.

Eighth grade, 7.62 per cent.

Ninth grade, 7.91 per cent.

Much of the senior high school geometry is of the demonstrative type. The tests are therefore mostly of this type.

J. H. Minnick [11] made a study concerning certain abilities fundamental to the study of geometry and concluded that the following four abilities are most important.

1. The ability to draw the figure for a theorem.
2. The ability to state correctly and accurately the hypothesis and conclusion of the theorem.

[10] Fifth Yearbook, Department of Superintendence, 1927, p. 205.

[11] "An Investigation of Certain Abilities Fundamental to the Study of Geometry," University of Pennsylvania, Philadelphia, 1918.

3. The ability to recall additional facts about a figure when one or more facts are given.
4. The ability to select from available facts those that are necessary for a proof, and to arrange them so as to arrive at the desired conclusion.

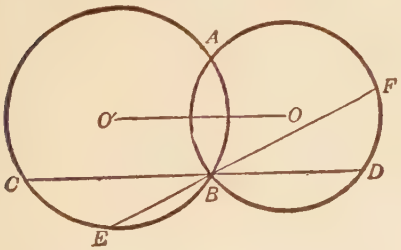
Following this study he constructed a series of tests[12] in geometry. Test *A* consists of a series of geometric exercises in which the pupil is told to "Read one of the exercises carefully, then in the space below draw the figure for the exercise, then read another exercise and draw the figure, continue in this manner. . . ."

Example: Test *A*

Draw the figure for the following proposition:
If two radii of a circle are perpendicular, and a tangent to the circle cuts these radii produced at points *A* and *B*, the other tangents drawn from *A* and *B* are parallel.

In Test *B* the theorem is stated, the figure is drawn, and the pupil is told to state below the exercise what is given and what is to be proved.

Example: Test *B*



I. State what is given and what is to be proved in the following proposition:
If two circles intersect, the common secant drawn through one of the points of intersection and parallel to the line of centers is greater than any other common secant drawn through that intersection.

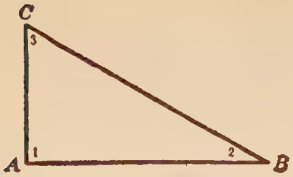
Given:

To Prove:

In Test *C* a figure is given, with certain minimum facts concerning the figure, and the pupil is asked to state as many more facts about the figure as he can.

[12] Published by the Public School Publishing Company.

Example: Test *C*



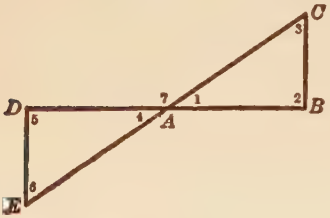
I.

Given: Triangle ABC, $\angle 1$ = a right angle, and $\angle 3$ = two times $\angle 2$.

State as many more facts about this figure as you can.

In Test *D* a figure is drawn, the hypothesis stated, and known facts given, and the pupil is requested to give complete proof of the exercise.

Example: Test *D*



I.

Given: AB = AD, ED is perpendicular to DB, CB is perpendicular to BD, and EC passes through A.

To prove that triangle ABC is congruent to triangle EAD.

Other known facts:

$\angle 6 = \angle 3$.

$\angle 1 = \angle 4$.

$\angle 4 + \angle 7 = 180°$.

$\angle 2 = \angle 5$.

ED is parallel to BC.

$AC - CB < AB$.

Proof:

This type of study and test is, of course, basic in character. Because of its subjectivity and complication in scoring, this test has not been so widely used as it might otherwise have been. The author has attempted through the scoring to make the test somewhat diagnostic.

Another test in plane geometry that is valuable is the Schorling-Sanford Achievement Test in Plane Geometry.[13]

[13] Bureau of Publications, Teachers College, Columbia University.

This test consists of five parts, each of which is made up of twelve elements. This test is based on the theory that, if the essence of geometry is demonstration and a consciousness of what constitutes a logical argument, we are interested (1) in a pupil's information about geometric facts, (2) in his ability to draw an inference from given data, (3) in his ability to judge the validity of a conclusion, and, as secondary matters, in his ability (4) to analyze a construction and (5) to use geometric facts in computation. Following are examples taken from the test.

I. Completion Sentences

[Instructions:] "In this part of the test, you will be given some sentences that are not complete. Read each one carefully, and think of a word which you might insert in each blank and so make the sentence the statement of a geometric fact. Write the word in the blank. Putting two or more words in the same blank counts against you." [Examples:]

1. A ..... is a line of unlimited length that touches a circle at only one point.
2. Parallel lines that cut off ..... segments on one transversal, cut off ..... segments on any other transversal.

II. Conclusions from Given Data

[Instructions:] "In this part of the test, you will be given geometric figures and certain facts relating to each. Think what fact you could prove about each figure on the basis of the given information and write a statement of this in the space beside the words "Prove that." Be sure to use each piece of information that is given to you." [Example:]

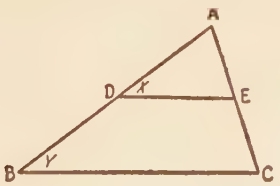


Fig. 1.

13-14 Use figure 1.
Given $\triangle ABC$ with $AD = DB$, and with DE drawn so that $\angle x = \angle y$.
Prove that (1) 13
Prove that (2) 14

III. Judging Conclusions

[Instructions:] "In this part of the test, you will be given a number of statements. Read each carefully. If it is true put a cross in the square marked 'True'. If the statement is false, put a cross in the square marked 'False.' But if it does not contain enough information to make sure that it is true or false, put a cross in the square marked 'You cannot tell.'" [Example:]

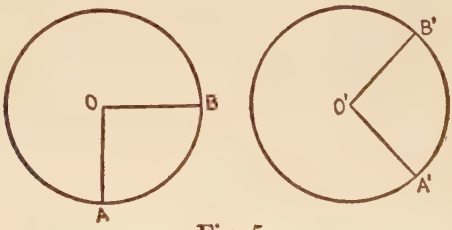


Fig. 5

	TRUE	FALSE	YOU CANNOT TELL	
25. In figure 5, if the two equal circles whose centers are O and O' have $\angle AOB = \angle A'O'B'$, then $\overset{\frown}{AB} = \overset{\frown}{A'B'}$.	☐	☒	☐	25

IV. Analyzing Constructions

[Instructions:] "In this part of the test, you will be given diagrams that illustrate certain geometric constructions. You are to study these diagrams and then you are to answer questions that are based on them. In some cases you are to complete a sentence that describes the diagram. You are not obliged to do this with one word as in Part I. Use as many

words as you really need, but make your answers as brief as you can and still make your meaning clear." [Example:]

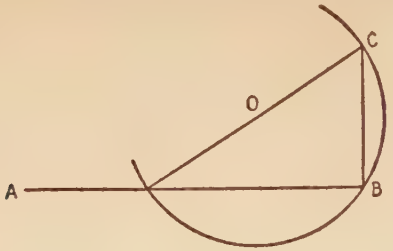


Fig. 12.

37. In figure 12, $\angle CBA$ is.....degrees. 37

V. Computation

[Instructions:] "In this part of the test, you will be given diagrams about which certain facts are known. You are to use these facts in computing such things as the length of a line, the size of an angle, etc., according to the questions that are asked. Read each question, study the figure upon which the question is based, and write the answer in the space beside the question." [Example:]

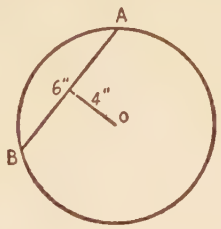


Fig. 18.

49. AB is a chord of the circle whose center is O (figure 18). AB is 6" long and it is 4" from O. What is the length of the radius of the circle?..... 49

The total time for the giving of this test is 52 minutes. Two forms of the test are available. Norms, given in Table 14, are for the end of the year.

TABLE 14.—END OF THE YEAR NORMS FOR THE SCHORLING-SANFORD GEOMETRY TEST

	Form A	*Form B*
75 percentile	37.0	36.8
Median	30.6	31.2
25 percentile	25.0	27.5

Another set of geometry tests, which is developed so as to be used at the conclusion of the study of the various books of geometry, is Hart's Geometry Tests.[14] These tests are newly published and have not as yet had wide enough use to have become completely standardized; nevertheless their construction is of a type that should be popular with geometry teachers who desire standard measures before the end of a whole year's work. Each of the five tests that make up this set is a test of recognition and application of the fundamental theorems of the book tested. Test I covers Book I, Test II covers Book II and so on, through Book V. The validity and reliability of these tests are not given by the publishers. Tentative norms are given. Only one form is available.

Tests in plane geometry have not as yet been developed to the place where they can be called diagnostic to any degree. However, several plane geometry tests are about ready to be put on the market. Some of these may prove to be diagnostic. The study of pupil difficulties in geometry has not advanced to a place that makes possible a discussion of remedial work. Progress in this phase of geometry teaching is held up, on the one hand, by the chaotic conditions of thinking concerning the place and value of geometry in the high school curriculum, and, on the other, by a lack of psychological study of pupil learning in geometry. An inventory test in both plane and solid geometry is much needed. Such an inventory test should cover the phases of geometry that are likely to have been covered in junior high school mathematics and those phases of

[14] D. C. Heath and Company, 1927.

algebra that the pupil needs to know in order to do his geometry work.[15]

A new test in solid geometry that has been developed in tentative form in such a manner as to be of some value in achievement testing and diagnosis is the Seattle Solid Geometry Test Series.[16] The following ten tests are to be used separately as they are needed: (1) and (2) lines and planes, (3) dihedral and polyhedral angles, (4) prisms and parallelopipeds, (5) pyramids, (6) cylinders and cones, (7) spheres, (8) spherical polygons and spherical areas, (9) spherical volume, (10) final examination. Each of these tests is made up of approximately fifty items. A cumulative score and a final test score provide an excellent basis for deciding the final grade. A chart for analyzing pupils' mistakes, used with each part, would be valuable as a basis for remedial teaching.

Following is a list of recommended tests in this field.

TESTS IN MATHEMATICS

ALGEBRA

Columbia Research Bureau Algebra Test. By Arthur S. Otis and Ben D. Wood. Published: 1927. Purpose: to provide a standardized final examination or college entrance test in first-year algebra. Range: high school and college. Time to give: 100 minutes. Time to score: 5 minutes. Standardized.
Publisher: World Book Co.
Price: $1.30 per package of 25, complete; specimen set, 20 cents, postpaid.

Dalman-Wentz Algebra Hurdles. By M. A. Dalman and E. K. Wentz. Published: 1926. Purpose: to measure students' work in first-year algebra. Range: grades 8 and 9. Time to give: varies. Time to score: varies. Not standardized.
Publisher: Public School Publishing Co.
Price: not available except in experimental form.

[15] A worthy attempt in this direction has been made by Leonard D. Haerter, "The Use of the Inventory Test in Plane Geometry," *The Mathematics Teacher*, March, 1926, pp. 147-54.

[16] Dvorak, Morgan, and Wait, "Seattle Solid Geometry Test Series," Seattle Public Schools, 1927.

Douglass Series of Diagnostic Tests for First-Year Algebra. By Harl Roy Douglass. Published: 1921; revised, 1924. Purpose: to measure achievement in first-year algebra. Range: grades 9 and 10. Time to give: Series A, 40 minutes for either form; Series B, 7 to 15 minutes. Time to score: short. Standardized.
Publisher: University Press, University of Oregon.
Price: Series A, $1.60 per 100 folders, complete; Series B, 4 cents per set of 7 tests, or $3.50 per 100; 3 cents per set of 5 tests (1-5), or $2.50 per 100; record sheets, 3 cents each; specimen set, 30 cents.

Hotz's First-Year Algebra Scales. By Henry G. Hotz. Published: 1918. Purpose: to indicate attainment, to measure progress, and to diagnose difficulties. Range: first-year algebra. Time to give: no time limit. Time to score: 30 seconds for each test of Series A; 60 seconds for each test of Series B. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: 80 cents per 100 copies of any scale, except the Graph Scale, which is $1.35 per 100; specimen set, 15 cents postpaid. "Manual of Directions for First-Year Algebra Scales," by Henry G. Hotz, 80 cents.

Illinois Standardized Algebra Tests. By W. S. Monroe and L. W. Williams. Published: 1920. Purpose: to measure speed and accuracy in finding the unknown "X." Range: secondary school. Time to give: 4 to 10 minutes for each of the 4 tests. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: $2.50 per 100; specimen set, 15 cents.

Rugg-Clark Standardized Tests in First-Year Algebra. By H. O. Rugg and J. R. Clark. Published: 1918. Purpose: to test pupils' achievement in algebra and to diagnose difficulties. Range: first-year algebra. Time to give: 4 minutes for each test. Time to score: short. Standardized.
Publisher: H. O. Rugg, 425 W. 123rd St., New York City.
Price: 8 cents per copy; specimen set, 12 cents.

Standard Tests of Progress in College Algebra. By Research University. In preparation. Purpose: to give comparative measures of progress in college algebra. Range: college. Time to give: about 50 minutes. Time to score: about 5 minutes. Standardized.
Publisher: not determined.
Price: not yet determined.

GEOMETRY

Columbia Research Bureau Test. Plane Geometry. By Herbert E. Hawkes and Ben D. Wood. Published: 1926. Purpose: to measure achievement in plane geometry. Time to give: 1 hour. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.20 per package of 25 examination booklets; specimen set, 25 cents.

Perry Geometry Tests. By Robert D. Perry. Published: 1927. Purpose: to measure achievement in geometry and to diagnose difficulties. Range: plane geometry classes. Time to give: 35 minutes. Time to score: 2 or 3 minutes. Standardized.
Publisher: Lafayette Publishing Co.
Price: 3 cents per copy; scoring keys, 3 cents each.

Renfrow Diagnostic Tests in Plane Geometry. By O. W. Renfrow. Published: 1926. Purpose: to measure pupils' knowledge of plane geometry. Range: plane geometry classes. Time to give: 45 to 60 minutes. Time to score: 45 minutes. Standardized.
Publisher: C. A. Gregory, University of Cincinnati.
Price: $1 per package of 25 test booklets, class record, and score cards. In quantities less than 25, 5 cents per copy and 5 cents extra for class record; specimen set, any form, postpaid, 10 cents, all forms, 35 cents.

Schorling-Sanford Achievement Test in Plane Geometry. By Raleigh Schorling and Vera Sanford. Published: 1923. Purpose: to measure achievement of pupils in plane geometry. Range: high school. Time to give: 52 minutes. Time to score: 2 to 3 minutes. Standardized.
Publisher: The Lincoln School of Teachers College, Columbia University.
Price: 10 cents per copy, Forms *A* or *B;* Manual of Directions, 25 cents.

Seattle Solid Geometry Test Series. By August Dvorak, M. E. Morgan, and W. T. Wait. Published: 1927. Purpose: to measure achievement in solid geometry. Range: high school. Time to give: 25 minutes for each part. Time to score: 1 minute for each part. Standardized.
Publisher: Seattle Public Schools, Seattle, Wash.
Price: 15 cents per set of 10 tests.

Webb Geometry Tests. By Paul E. Webb. Published: 1926. Purpose: to test pupils' ability in geometry. Range: students of high school geometry. Time to give: 45 minutes. Time to score: short. Standardized.

Publisher: Public School Publishing Co.
Price: 75 cents per package of 25 tests; specimen set, 15 cents.

COMMERCIAL ARITHMETIC

Kinney Commercial Arithmetic Scales. By L. B. Kinney. Published: 1926. Purpose: to test pupils' ability to solve problems similar to the problems found in the business world. Range: high school. Time to give: Test *A,* Part I, 20 minutes; Part II, 10 minutes; Tests *B* and *C,* no information. Time to score: no information. Standardized.
Publisher: Public School Publishing Co.
Price: $1 per package of 25 tests; specimen set, 10 cents.

TRIGONOMETRY

Standard Tests of Progress in Trigonometry. By Research University. In preparation. Purpose: to give comparative measures of progress in trigonometry. Range: college. Time to give: about 50 minutes. Time to score: about 5 minutes. Standardized.
Price: not yet determined.

BIBLIOGRAPHY

Barton, W. A. "Effect of Group Activity and Individual Effort in Developing Ability to Solve Problems in First-Year Algebra," *Educational Administration and Supervision,* Vol. 12: pp. 512-18, November, 1926.

Breslich, E. R. "The Needs of the Superior Pupil in the Study of Mathematics," *23rd Yearbook National Society for the Study of Education,* Part I: pp. 372-78.

Deam, Thomas M. "Diagnostic Algebra Tests and Remedial Measures," *School Review,* Vol. 31: pp. 376-79, May, 1923.

Dickinson, E. L. and Ruch, G. M. "Analysis of Certain Difficulties in Factoring Algebra," *Journal of Educational Psychology,* Vol. 16: pp. 323-28, May, 1925.

Hobbs, James B. "Results from Giving the Hotz First-Year Algebra Scale Tests to a Six-Eight Months Group," *School and Society,* Vol. 12: pp. 353-54, October 16, 1920.

Hotz, Henry Gustave. "First-Year Algebra Scales," *Contributions to Education,* Teachers College, Columbia University, No. 99, 1918.

Lohr, V. C. "Study of Mathematical Abilities, Powers, and Skills as Shown by Certain Classes in Physical Science," *School Science and Mathematics,* Vol. 25: pp. 834-44, November, 1925.

Monroe, W. S. "Measurements of Certain Algebraical Abilities," *School and Society,* Vol. 1: pp. 393-95, March 13, 1915.

Rugg, H. O. "The Experimental Determination of Standards in First Year Algebra," *School Review,* Vol. 24: pp. 37-66, January, 1916.

Rugg, H. O. and Clark, J. R. "Standardized Tests in the Improvement of Teaching in First-Year Algebra," *School Review,* Vol. 25: pp. 113-32, February, 1917.

Rugg, H. O. and Clark, J. R. "The Improvement in Ability in the Use of the Formal Operations of Algebra by Means of Formal Practice Exercises," *School Review,* Vol. 25: pp. 546-54, October, 1917.

Stockard, L. V. and Bell, J. C. "A Preliminary Study of the Measurement of Abilities in Geometry," *Journal of Educational Psychology,* Vol. 7: pp. 567-80, December, 1916.

Stone, C. A. "Construction of a Test to Measure Mathematical Ability," *School Science and Mathematics,* Vol. 26: pp. 824-32, November, 1926.

Webb, P. E. "A Study of Geometric Abilities Among Boys and Girls of Equal Mental Abilities," *Journal of Educational Research,* Vol. 15: pp. 256-62, April, 1927.

Wentz, Kate. "Hurdle Tests in Algebra," *School Science and Mathematics,* Vol. 25: pp. 132-44, February, 1925.

Yingling, R. W. "Diagnosis and Training in Advanced High School Algebra," *School Science and Mathematics,* Vol. 26: pp. 729-34, October, 1926.

CHAPTER VII

Language

For the purpose of this chapter language is considered in terms of language usage, composition, grammar, and literature.

There is, of course, one chief aim in language teaching in both elementary and secondary school instruction; that is the development of the art of communication in both oral and written use of language. Certain other life objectives, like formation of character or training for citizenship, may be guiding influences, but they are not the primary objectives to be attained through language instruction. Language as a tool of thinking should doubtless receive more emphasis than it usually does receive in the statement of aims and objectives or in the practice of teaching language in the schoolroom.

This general aim of communication is more concretely and definitely stated in the following specific aims:[1]

1. To teach the child to communicate freely, effectively, and correctly in speech or in writing in those life situations which demand communication.
2. To enlarge the child's vocabulary and to stimulate effective choice of words.
3. To teach the child to collect, organize, and present materials for oral and written discussion.
4. To make automatic certain habits for effective oral and written speech.
5. To develop taste for good language and foster ideals of proper speech and style.
6. To teach the child to pass critical judgment on his own speech and writing.

[1] Clifford Woody, *Language Aims, Suggestions, and Activities for Their Realization.* Bureau of Educational Reference and Research, University of Michigan, Ann Arbor, Mich.

The content of a standard test in language, and the grades for which it is suited, are dependent upon the grade placement in the course of study of the various phases of the material tested. There seems to be less agreement than would be expected in respect to the grade placement of the various phases of the language course of study. Although grammar is a pretty well defined phase of language work, Miner [2] found that there is far from complete agreement concerning the grades in which it should be taught. This was revealed by his analysis of 88 courses of study in language.

TABLE 15.—VARIOUS GRADES IN WHICH TECHNICAL GRAMMAR IS TAUGHT (MINER)

Grade	*Schools*
III	1
IV	2
V	30
VI	56
VII	58
VIII	49

The preceding table shows the number of schools in Miner's analysis teaching technical grammar and the grades in which the subject is taught. From this tabulation it is noticed that, while the largest number of schools teach grammar in grade seven, there are a good many that teach it as early as the fifth grade. However, it seems that the seventh grade is the grade in which the largest number of school systems teach grammar.

AN ANALYSIS OF ABILITIES IN LANGUAGE

Composite language ability increases with age and grade. This fact may easily be proved by observation or by scientific studies. Much of the development of language usage, both written and oral, is the result of the formation and continua-

[2] "The Present Status of Language Teaching in the Elementary Schools as Revealed by Eighty-eight Courses of Study," M.A. thesis, Indiana University, Bloomington, Indiana, 1925.

tion of specific habits or bonds. "Good English is a matter of habit" is a phrase often on the blackboard or on a placard in the classroom. Perhaps the phrase should read, "Good language usage is a matter of habits." An analysis, then, of language usage requires that we study in children the formation of habits. Habits are formed by conscious and unconscious imitation and repetition. Meaningful experiences well associated, accompanied by a desire and the opportunity for the child to tell about these experiences, develop the content out of which the habits of language grow. This communication to others of these associated experiences or desires of the child may be correctly or incorrectly made. The correctness or incorrectness of this communication becomes the form side of the language. Habit formation and bond development play an important part in developing good form. In the experiences and needs of children and in their communication concerning their experiences and needs, language grows in content and form, each somewhat limiting the other. From the time the child learns to write until he completes his school work, language usage is objective in form and can consequently be measured. Correct use of words, sentence structure, punctuation, spelling, capitalization can be objectively measured to a marked degree.

Formal grammar consists of the technical rules of language communication. It is pretty well agreed that these technical rules need not be taught as rules until after the child has played the game of language for some time. "You cannot play baseball until you have committed all the rules" is not a true statement. The teacher of games goes on the theory that much of the game can be played at first with little knowledge of the rules. The child is told how the game is played; he is taught by the "master of the game" what to do and how to do it. A knowledge of the rules is picked up by playing the game, but the learning of the technical rules so that they can be quoted and applied is reserved for a much later time and usually for the expert player and the umpire. The same may

be said of the relation between grammar and language usage. The rules should grow with the game of usage. Study of these rules may well wait until the junior high school level at least. It is a question whether the location of formal grammar even on this level is psychologically correct from the child's point of view or propædeutic in the scheme of subjects. At least, "sensorimotor bonds come first, and high-level rules later." [3]

Measurement in language is at present deficient in several ways.

(1) There is no satisfactory method of rating or scoring or objectively measuring oral ability or achievement in language. This is, however, an important phase of language work. A scale of oral composition may later be developed by means of phonograph records. An oral composition score card might be used to aid in judging oral compositions. This would help to call attention to the items that are important. Such a score card should also indicate the relative importance of various phases of oral English.

(2) Language tests are based on the mechanical and formal aspects of language. They do not give reliable measures of function and content.

(3) An analysis of composite language ability and the development of it has not been carried far enough in certain phases to allow for specific and diagnostic measurement.

(4) There is a dearth of composite measures for use in the lower grades of even those factors that are measurable.

(5) There is assumed a relation between "doing" and "knowing," and "correcting" and "using," that does not necessarily exist.

The tests in language are usually listed under the heads of Language Usage, Grammar, Composition, and composite measures of various phases of language.

[3] D. W. LaRue, *The Child's Mind and the Common Branches* (The Macmillan Company, 1924), p. 247.

LANGUAGE USAGE

The tests in language are based largely upon the most common errors made by children. Such a list of errors has been made up as a result of scientific studies of children's written and spoken errors undertaken by such writers as Wilson, Charters, Clapp, and others. Several tests will be discussed in some detail.

The Wilson Language Error Test[4] is composed of three stories—Story A, Story B, Story C—that are approximately equal in difficulty. Each of these stories contains 28 common and typical errors made by children in language usage. The test is primarily a measure of the objective "to make automatic certain habits for effective . . . written speech." The reliability for a single grade for a single test is approximately .75, for several grades, .90. The validity as represented by comparison with teachers' marks is approximately .45 with 42 cases.

The manual of instructions is complete with the exception of data on reliability or validity. The key showing the errors accompanies the package of tests. The directions for scoring are specific and definite. Norms are available for grade and age.

Following are the instructions for administering the test to children of upper third to twelfth grades. These instructions assume that the information blanks of the test such as name, age, etc., have been filled in.

"Now I will read the directions on the first page. This is a test to see if you can correct the mistakes that a pupil has made in writing a story. Here is a very short sample story. See if you can find the mistakes in it:

"Sample story: Willie come to visit us. He is only six years old. He stayed a hour. He has went home. I like to Willie.

"Here is the same story corrected:

[4] Published by the World Book Company.

"Sample story: Willie *came* to visit us. He is only six years old. He stayed *an* hour. He has *gone* home. I like Willie."

(The incorrect and the corrected samples are shown.)

"That is the way you are to do in this test. There are three stories in this folder, but you will take only the first one unless I tell you otherwise. You are to correct all the mistakes in that story just as has been done in the sample. Draw a line through each wrong word and write the correct word above it. Work as fast as you can, but be very careful to correct every mistake. You will be given time enough to finish unless you are very slow. Do you understand just what you are to do?"

(Answer any questions as to the method of taking the test.)

If only Story A is to be given, say: "Now turn to Story A on the next page and fold your paper back so that only one page is in sight. When you finish this story, close your paper and leave it on your desk with the first page up. Just take one story. Begin."

Following is a part of the scored paper of a pupil, S.L.D., of the 6B grade. One of the instructions given to the scorers is to indicate each error corrected by a check mark (√) opposite it in the *right-hand* margin. Underscore each error not corrected and place a check mark opposite it in the *left-hand* margin. In this manner the total check marks on the right indicate the total errors corrected, while those on the left indicate the errors not corrected.

STORY A

SATURDAY MORNING

		at
√	Saturday morning is a busy time to are house.	√
		Dorothy
√	A feller has a good chance to work. Me and	√
	I	
	Dorothy divide the tasks between us. Then we	√
	race to see who will finish first. Last Saturday	
	took	
	I taken the breakfast dishes as one of my tasks.	√

V I am especial fond of washing dishes. You

V should have saw me work. I wanted to get

A study of the preceding paper would give a picture of the types of errors recognized and corrected by the pupil and those evidently not recognized or at least not corrected. From this test paper these things can be noted: (1) This boy seems to recognize and correct the improper use of *to, too,* and *two.* This ability indicates an understanding of, and an accompanying proper usage of, these terms. (2) He does not see and correct the use of *saw* for *seen,* nor *seen* for *saw.* Note lines 4, 7, 14.[5] This diagnostic analysis may be made of a whole class and the results be carried directly to the pupil and the class by such a chart as follows (Table 16), based on the tests of twenty-nine 6A pupils in the same school.

An analysis of certain errors or habits of each pupil both in terms of score and in terms of his particular difficulty is given by the table on pages 139 and 140. The difficulties of the class as a whole and the standing of the class as a whole can also be ascertained from a study of this table.

Reference to the table shows that certain types of errors were not corrected by a rather large per cent of the group. "I sure appreciated the doughnuts" was not corrected by a single child of this grade. "I ain't got no marbles" was corrected by almost all of the pupils of the same grade. A glance at the table will show that certain errors were corrected by a very few of the class. Such an analysis points to specific needs of the class as well as to its accomplishments.

The total number of errors corrected by the pupils varied from only 11, the number corrected by pupil A.M.W., to 25 of the possible 28, corrected by pupil R.B.C. The median score or median number corrected by the class was 22. The norms are given for the various grades on page 141.

[5] Not all of the lines in the test are here reproduced.

TABLE 16.—A DIAGNOSTIC ANALYSIS OF THE RESULTS OF AN APPLICATION OF THE WILSON LANGUAGE ERROR TEST, STORY A, TO 6A CLASS IN THE XYZ, INDIANA, SCHOOLS, 1927

Name	*lo*	*are*	*feller*	*me*	*Dorothy*	*taken*	*especially*	*saw*	*so as*	*be*	*too*	*couldn't hardly*	*did*	*sure*	*real*	*ain't*	*got*	*no*
A.J.A.							x				x			x				
T.M.B.		x[6]										x	x	x	x			
E.E.B.												x		x	x			
V.L.C.		x		x	x		x	x			x	x	x	x	x			
V.A.C.		x					x							x	x			
L.S.C.														x	x			
A.D.C.							x	x						x	x			
R.V.C.	x		x		x		x	x				x	x	x			x	x
R.B.C.												x		x				
L.K.F.						x								x				
R.O.F.		x	x									x		x	x			
L.P.G.						x						x		x	x			
H.J.H.		x	x				x	x				x		x	x			
M.W.H.		x					x							x	x			
C.E.J.							x	x	x			x	x	x	x			
D.J.K.							x					x		x	x			
H.R.M.				x			x	x		x	x	x	x	x	x		x	x
C.R.M.								x					x	x				
J.L.P.		x	x				x					x	x	x	x			
M.L.P.		x	x									x		x	x			
H.E.R.			x					x				x		x	x			
D.E.R.		x	x			x	x			x			x	x	x			
M.G.R.			x					x				x		x	x			
B.D.S.		x	x	x	x		x	x	x		x	x	x	x	x			
M.E.S.			x				x					x		x	x			
P.E.T.			x									x		x	x			
D.L.V.							x					x		x	x			
A.M.W.	x	x	x				x	x	x		x	x	x	x	x			
L.M.W.							x							x	x			
TOT'L FAILURE TO CORRECT	2	11	12	3	3	3	17	11	3	2	5	20	10	29	24	0	2	2
TOTALS IN PERCENTS	7	38	41	10	10	10	59	38	10	7	17	69	35	100	82	0	7	7

[6] X on the chart represents the failure of a pupil to correct the error.

[Chart continues on page 140.]

[Table 16 continued.]

Name	comes	give	to	was	seen	et	them	was	couldn't	hopes	Score: Total No. corrected	Score: Total No. not corrected
A.J.A.		x	x	x							22	6
T.M.B.	x			x	x						20	8
E.E.B.			x								24	4
V.L.C.		x	x	x	x			x	x		12	16
V.A.C.				x							23	5
L.S.C.			x	x							24	4
A.D.C.				x							23	5
R.V.C.			x	x		x					15	13
R.B.C.				x							25	3
L.K.F.	x		x	x							23	5
R.O.F.					x						22	6
L.P.G.											24	4
H.J.H.				x	x			x			18	10
M.W.H.		x		x							22	6
C.E.J.	x		x	x							18	10
D.J.K.			x	x							22	6
H.R.M.			x	x	x		x	x	x		11	17
C.R.M.			x	x							23	5
J.L.P.			x	x			x	x	x		16	12
M.L.P.		x	x	x							20	8
H.E.R.			x	x							21	7
D.E.R.		x			x		x	x	x		15	13
M.G.R.			x	x							21	7
B.D.S.			x	x	x			x	x		11	17
M.E.S.				x							22	6
P.E.T.			x								23	5
D.L.V.			x	x							22	6
A.M.W.		x	x	x	x	x			x		11	17
L.M.W.				x	x						23	5
TOT'L FAILURE TO CORRECT	3	6	18	24	9	2	3	6	6	0	..	236
TOTALS IN PERCENTS	10	21	62	82	31	7	10	21	21	0	..	29

Estimated Median Scores for Grades 3 to 12

Grade	3	4	5	6	7	8	9	10	11	12
Median Score	7	11	14	17	20	22	23	24	25	26

After finding the types of errors made by children, the next task is to find the cause of the errors. Are such errors due to a lack of knowledge of what is the correct way to use the word, or of the distinction in the meaning of the words, as "to, too, two"? Is it the likeness of sounds of words, as "at our (are) house," that confuses children? Is the error due to habits of imitation of wrong usage which are of such long standing that they are difficult to break and are these errors consequently not noticeable when seen in incorrect form in copy to be corrected? Diagnostic testing inevitably leads to two questions closely related to testing: How do children think? and What is the best method of teaching?

How do children think concerning "to, too, and two"? Do they notice that the words are different? In oral speech this distinction does not have to be made; therefore the problem is one of written communication only. Some children distinguish by association with like usage in another situation, others by the spelling and use, others by meaning. Re-teaching, of course, is not unlike first teaching except that it may be more specific and emphatic. It may require much explanation with meaningful illustrations drawn from many situations. This explanation should be followed by drill on the right form for the purpose of fixing the right habit. It is necessary to enlist the interests of the pupils and to create a desire on their part to do away with these difficulties. The teacher of the class whose chart is shown here made a large chart somewhat like the one shown in Table 16 and placed it in the school room. The X was a derogatory mark on the pupils' records, and they desired to remove it. To remove the mark, the pupil must convince the class and the teacher that the error would never be made any more. When the pupil accomplished this feat, he

had a right to erase the X after his name in the column of the particular error marked against him on the sheet. In this manner he could also raise his score and eventually have a clear sheet. Drill material that can be individually applied is not available in this line of work and must be provided by teachers. After remedial work has been done, a duplicate form, Story B or C, can be used to determine progress.

The adverse criticisms of the Wilson Language Error Test are that (1) it affords a rather limited sampling of language usage situations; (2) it measures only one phase of the results of language instruction, and, therefore, is not a safe criterion upon which to base a judgment as to the progress of the class in language work; (3) in the scoring of the test no account is taken of mistakes made by the pupil in changing correct expressions into incorrect ones. It is suggested that errors in this latter class are indicative of a pupil's lack of knowledge of language usage and should be considered, at least in the diagnostic use of this test.

The Charters' Diagnostic Language Tests[7] are based on studies of language errors of children. These tests measure somewhat the same objective measured by the Wilson Language Error Test. There are three of the tests—one each in verbs, pronouns, and miscellaneous. Each test has two forms. The tests are best used above the fifth grade, although they may be used from grades three to twelve.

The manual of directions is not complete enough. The scoring is objective. The tests are easily and quickly scored. Norms by 25 percentile, median, and 75 percentile are available by grades. Each test is made up of 40 sentences. The child is directed to "read the sentence carefully and see if it is right. If it is right you are to make a cross on the dotted line below the sentence." (Example is given to pupils.) "If the sentence is not right you are to put the correct word or words on the

[7] Published by the Public School Publishing Company.

dotted line below it." (Example is given.) Some examples are given below from the test in verbs.

14. I haven't wrote it yet.
15. He done his work yesterday.
16. Once he give me a ride.
17. You must lay down every day.
18. Foxes have often ate chickens.
31. Where does George set?
32. They have took the dog away.
33. Mr. Fox has came.
34. When he entered the bear begin to growl.
35. Each of them were afraid.

This type of test offers an easy method of diagnosis. The errors used in the test are typical of pupils' work, and the number used in the test gives a fair sampling of the most common errors.

The test easily lends itself to coaching preparatory to taking another form of the test to measure progress. A teacher told a class, in discussing with them the first test, that all the examples were wrong except the first one and should have been corrected. Not so long afterwards, when a second form was given, the children worked on the theory that most of the sentences were wrong and it was their work to find out what was the matter with them. Such knowledge on the part of the pupils resulted in a gain not at all consistent with the habits of usage. The test does not so nearly represent classroom product as the composition type like the Wilson test. Willing [8] found that correction of printed matter does not correlate as high with ability to write correct English as might be expected. In view of this fact the main value of the Charters Language Usage Test lies in its diagnostic and remedial functions.

The Kirby Grammar Test attempts to measure (1) the ability to choose correct usage when two forms are given and (2) the ability to identify that choice in terms of the rule of grammar that makes it the correct choice. The content in usage is based on studies of most common typical errors made by children. The test measures most nearly objectives 4 and 6

[8] *Valid Diagnosis in High School Composition*, Teachers College Contributions to Education, No. 230, 1926.

found on page 132. The test is divided into five sections: (1) use of personal pronouns, (2) possessive pronouns and contractions, (3) commonly misused verbs, (4) common verb errors, (5) miscellaneous. The pupil is tested as to his ability to choose the correct usage and then from a list of rules to choose the rule that applies to the sentence corrected.

The manual of directions is not complete. Instructions for scoring are fair. The test is easily and quickly scored, and is objective. Norms are given by grades, VII to XII, for both usage and principles. The reliability is given in Table 17.

TABLE 17.—DATA ON THE RELIABILITY OF THE KIRBY GRAMMAR TEST

(Adapted from Ruch & Stoddard[9])

Principles	*No. of Cases*	*Nature of Group*
.87	30	Grade VII
.90	18	Grade VIII
.91	128	Grades VII-XII
Sentences		
.60	32	Grade VII
.43	18	Grade VIII
.70	136	Grades VII-XII

The table may be read as follows: The reliability of the part of the test on "Principles" has a coefficient of reliability of .87 based on 30 cases in Grade VII, etc.

The following is a section from the instructions to those taking the test, together with a section from the test itself.

Read the first sentence in Section I; then mark out the incorrect form. Read the rules in Section I, until you find the one that applies to this first sentence. Place the letter of this rule in the square preceding the first sentence. Proceed with the second sentence in the same way, and so on until you finish Section I. Then continue with Section II, and so on through five sections. If you are unable to find the rule for any sentence, go on to the next sentence, but do your best to find the correct rule for each sentence.

[9] *Tests and Measurements in High School Instruction* (World Book Company, 1927), p. 100.

SECTION III [10]

Sentences

☐ 1. Either the boy or his sister always (goes) (go).
☐ 2. Bruno (lay) (lie) there in the corner.
☐ 3. We had (come) (came) on ahead.
☐ 4. My brother (lay) (laid) on the grass.
☐ 5. The child was (laying) (lying) on the grass.
☐ 6. They (have) (have got) to try harder.
☐ 7. There (were) (was) several ways to try.
☐ 8. They were (lying) (laying) the foundation yesterday.

Rules or Principles

a. In sentences beginning with "There" followed by "is," "are," "was," "were" and other forms of the verb "be," the verb agrees with the subject which follows.
b. Singular subjects joined by "or" and "nor" take a singular verb.
c. The imperative form of the verb is the first principal part.
d. Simple past time is expressed by the second principal part of the verb.
e. When making a verb form with "has," "have" and "had" use the third principal part of the verb.
f. "Lying" is a form of the intransitive verb "lie."
g. Avoid the use of "got" with forms of the verb "have."
h. "Laying" is a form of the transitive verb "lay."

COMPOSITION TESTS

There are two commonly used ways of measuring compositions. One is by measuring the form and mechanics of the composition; the other is by developing a standard or a scale of composition with which an individual composition is compared on a general or some particular basis. It is relatively easy to measure the mechanics and form side of composition and very difficult to measure the "content-as-expressed" objectively. Many factors enter to make it difficult to measure accurately this latter phase of achievement in composition. The abilities necessary in writing a composition are many, and the conditions of writing are hard to standardize properly. Composition scales measure to a limited degree all the objectives of language that have to do with written expression.

[10] Published by the Extension Division, University of Iowa.

A widely used composition scale that may be used for grades IV-IX is the Willing Composition Scale. This scale was developed through the use of children's compositions. It measures both quality, i.e. story value and form—the latter in terms of mistakes.

The composition scored 20 represents the poorest effort in story value and the most mistakes, and the one scored 90 the best effort in story value and no mistakes. The compositions in between represent equal steps of quality from poorest to best in story and in diminishing number of errors. The lowest one in quality of story value is designated as 20; they continue upward by steps of ten to the highest, which is 90.

SCALE FOR MEASURING WRITTEN COMPOSITION [11]

(Devised by M. H. Willing)

20

Deron the summer I got kicked and sprain my arm. And I was in bed of wheeks And it happing up to Washtion Park I was going to catch some fish. And I was so happy when I got the banged of I will nevery try that stunt againg.

Number of mistakes in spelling, punctuation, and syntax per hundred words, 30.

50

One time mother and father were going to take sister and I for a long ride thanksgiving. We had to go 60 miles to get there, When sister and I herd about it we were very glad. It was a very cold trip. We four all went in a one seated automobile. Dady drove and mother held me and sister sat on the top the top was down. Mother could not hold sister for she was two heavy. When we got there they had a hot fire ready for us and a goose dinner. We were there over night. In the morning it was hot out. This was on a farm. Sister and I got to go horseback riding. It was lots of funs. They had children. The children were very nice. Our trip home was very cold. When we got home it had snod.

Number of mistakes in spelling, punctuation, and syntax per hundred words, 14.

[11] Published by the Public School Publishing Company.

90

The most exciting experience of my life happened when I was five years of age. I was riding my tricycle on the top of our high terrace. Beside the curbing below, stood a vegetable wagon and a horse. Suddenly I got too near the top of the terrace. The front wheel of my tricycle slipped over and down I went, lickety-split, under the horse standing by the curbing. I had quite a high tricycle and the handlebars scraped the horse's stomach, making him kick and plunge in a very alarming manner. I was directly under him during this, but finally rolled over out of his way and scrambled up. I looked at my hands! Most of the first finger and part of the thumb of my left hand were missing. The horse had stepped on them. I had endured no sensation of pain before this, but now my mangled hand began to hurt terribly. I was hurried to the hospital and operated on, and now you would hardly notice one of my fingers is missing. I certainly have good cause to congratulate myself on my good fortune in escaping with as little injury to myself as I did, for I might have been terribly mangled in my head or body.

Number of mistakes in spelling, punctuation, and syntax per hundred words, 0.

In administering the Willing Composition Test, the examiner tells the pupils to write a story "about some exciting experience that you have had or about something very interesting that has happened to you." Suggestions of types of topics about which to write are written at the top of the paper upon which the pupil writes the composition. About twenty minutes are allowed for the writing. The pupil's paper is rated for story value first by neglecting all errors of form and keeping in mind only the value of the story which the pupil is telling. This story value is determined by comparison of the pupil's story with the scale. The scorer may give values between the units of the scale if he so desires. The compositions are then rated for form value by the careful marking of all errors in grammar, punctuation, capitalization, and spelling. The total number of errors of all types mentioned is then multiplied by 100 and divided by the total number of words in the composition. This gives a form value. The story value and form value may be distributed on a chart like the one given in

Figure 9. On this chart the heavy lines represent the norm for the grade. Those pupils who are in the upper right-hand

Make a Separate Sheet for Each Grade (Yearly or Half-Yearly)

City Bloomington State Ind.
Date March, 1927 School X
Teacher Brown Grade 5A

	STORY VALUE								
Errors per Hundred Words	20	30	40	50	60	70	80	90	*Distribution for Errors*
0 to 2.9									
3 to 5.9		1				3	1		5
6 to 8.9			1		1	1			3
9 to 11.9									
12 to 14.9			2	1					3
15 to 17.9		1				2			3
18 to 20.9			1						1
21 to 23.9		1					1		2
24 to 26.9									
27 to 29.9									
Above 30									
Distribution for "story value"		3	4	1	1	6	2		17

MEDIAN SCORES

GRADE	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*
Story Value Score	41.5	58.7	68.1	74.0	76.6	79.0
Form Value Errors	18.5	10.7	6.8	5.8	4.5	4.4

FIG. 9.—Class Record Sheet for Willing's "Scale for Measuring Written Composition." [12]

[12] Published by the Public School Publishing Company.

corner are above the norm both in story and form, while those who are in the lower left-hand corner are below both in story and form value. Those who are in the lower right-hand corner are above in story value but low in form. Those in the upper left corner are above the norm in form and below in story value. By a glance at the chart it is possible to see the condition of the class as a whole. This particular class is somewhat low in both story and form value. Further diagnosis of the form phase can be made by making a diagnosis by listing the types of errors made. By making a diagnostic analysis of these errors similar to the one used with the Wilson Error Test (Table 16) a diagnostic chart is obtained that will point the way toward remedial work.

Composition mechanics or form may be measured by tests other than the composition scale. This type of measuring is highly objective and can be easily used for diagnostic purposes.

The Pressey Diagnostic Tests in English Composition [13] are a series of four separate tests—capitalization, punctuation, grammar, and sentence structure. Each test has two forms. These tests were developed after a careful study of the purposes, uses of, and mistakes in, these phases of composition. Because these tests are well arranged and are accompanied with a specific statement of what they are intended for, it is easy to use them. There are three manuals that may be used with these tests. One contains the plan of the test, the directions for giving, scoring, recording scores, and norms by grades. Statements of reliability and validity are not given. Another is a student's guide to correctness in written work; and still another is the teacher's manual. These last two serve as guides to remedial work. The test maker may well consider it a part of his task to develop such remedial manuals to go with diagnostic tests. While the teacher will not be able to apply all the remedial work, she will find such manuals very suggestive.

[13] Published by the Public School Publishing Company.

Capitalization

In this test there are twenty-eight sentences. The pupil is told that "the sentences are correct in every way except that they lack all capitalization—aside from capitals at the beginning of each sentence. You are to draw a line under each letter that should be a capital. . . ." The rules tested for and the sentences that exemplify each rule are given here. These rules, of course, do not appear in the test, but the sentences are taken from the test itself.

(1) Capitalize the names of persons, with their titles; however, do not capitalize titles when they are not a part of a name.

1. There go tom black and mary smith.

(2) Capitalize the names of countries, states, cities, streets, buildings, mountains, rivers, oceans, or any word designating a particular location or part of the world; however, do not capitalize the points of the compass.

7. The rhine flows from the alps to the baltic sea.

(3) Capitalize the names of business firms, schools, societies, or other organizations.

9. He graduated from east high school and then went to yale.

(4) Capitalize words derived from the names of countries, places, organizations, or persons.

16. The candidate spoke with rooseveltian vigor.

(5) Capitalize the first word and all other important words in titles (and sub-titles and headings), of themes, magazine articles, poems, books of laws or governmental documents, and the trade names of commercial products.

19. The declaration of independence was then read.

(6) Capitalize the days of the week, the months of the year, and holidays and church days; however, do not capitalize the seasons.

24. The children's favorite holidays are christmas and thanksgiving.

(7) Capitalize the first word of every sentence, the first word of every line of poetry, and the first word of a direct quotation; however, if the quotation begins in the middle of a sentence (or is indirect) do not use the capital.

28. "Oh, what a tangled web we weave
when first we practice to deceive!"

Punctuation and Grammar

In the same manner the ability in punctuation, grammar, and sentence structure is tested. Examples follow.

(1) Full stops: Use a question mark after a direct question (not after an indirect question). Use an exclamation point after a sentence, exclamation, or interjection to show strong emotion or surprise. At the end of all other sentences use a period. Use a period also after abbreviations and initials.

18. Quick Come What a sight.

A portion of the test in grammar tests the following principles. The directions for the test are also given.

(1) Be careful about the tense forms of verbs. (a) Do not leave off the *ed* of past tense and participal forms of regular verbs; this mistake is most common with *ask, seem,* and *use.* (b) Be certain of the correct past tense and past participle of the following irregular verbs: *come, see, run, do, go, give, bear, lead, grow, hear, ride, drink, burst.* (c) Remember that *drown, climb, drag,* and *attack* are regular.

Directions: Below there are some groups of four sentences. In each group one sentence, and only one, is incorrect—contains something which is not good English. You are to find the wrong sentence in each group and put a cross in the square before it. If you are not certain which sentence is incorrect, mark the one which seems to you least satisfactory; be sure to mark one, and only one, sentence in every group. Remember

also to mark in the square; do not put a cross halfway between two sentences. Any doubtful marks will be counted against you.

11. ☐ The time seemed to pass very swiftly.
☐ She folded, sealed, and mailed the letter.
☐ He ask me the way to the station.
☐ We have used that sewing machine for years.

12. ☐ He almost drowned while trying to swim across the river.
☐ John clumb the tree to the limb where the kitten was.
☐ The police dragged the river for the body.
☐ At recess we attacked the snow fort.

13. ☐ I saw Harold at school this morning.
☐ Mary did all the things that you wished to have done.
☐ Arthur came over to see you yesterday.
☐ He run until he was out of breath.

14. ☐ Our heater bursted and flooded the cellar.
☐ She heard all that you said.
☐ The path led to the top of the hill.
☐ The corn grew fast during the hot spell.

15. ☐ He had drunk all the lemonade.
☐ Mother has given Harry a watch for his birthday.
☐ Helen was borned on Christmas day.
☐ I have never ridden over worse roads.

Guiler,[14] in making a study of thirty-nine college students that were failing or nearly failing in English composition, used as one of the diagnostic tests, the Pressey Punctuation Test. He listed the following twenty-four abilities measured by the test:

Ability to Use

1. A comma to separate words in a series.
2. A comma to separate parts of an address.
3. A comma to set off a direct quotation.
4. A comma to separate parts of a date.
5. A comma to set off slightly parenthetical or inserted expressions.

[14] "Diagnosing Shortcomings in English Composition," *Journal of Educational Research,* Vol. 14, No. 2 (September, 1926), p. 112.

6. A comma to separate clauses joined by "and," "but," "for," "as," etc.
7. A comma to set off obviously added elements at the end of a sentence.
8. A comma to set off an appositive word, phrase, or clause.
9. A comma to set off a clearly introductory expression.
10. A comma to set off an adverb used independently.
11. A semicolon to separate clauses of a compound sentence not joined by a conjunction.
12. A semicolon to separate parts of a compound sentence containing commas.
13. A period after an abbreviation.
14. An apostrophe to indicate the omission of a letter or letters in a word.
15. An apostrophe to show possession.
16. Quotation marks to enclose a direct quotation.
17. Quotation marks to enclose each part of a divided quotation.
18. Quotation marks to indicate the title of a theme, story, or song.
19. Quotation marks to call attention to words used with some special meaning.
20. A dash to indicate a marked break in the progress of thought in a sentence.
21. A question mark after a direct question.
22. A colon after an expression that serves as a formal introduction to something that follows.
23. A colon between hours and minutes in a statement of time.
24. An exclamation point to indicate strong feeling or emotion.

A diagnostic chart showing an analysis of pupil difficulties in terms of the abilities listed is shown in Table 18.

It will be noted from this table that some inabilities were manifested by most of the pupils, while others were not. Almost all showed inability to use a dash to indicate a marked break in the progress of thought in a sentence. No one manifested inability to use a comma to set off an adverb used independently. This type of analysis makes possible directed reteaching.

TABLE 18.—MEASUREMENT OF PUNCTUATION ABILITIES (GUILER)

Students' Numbers	*Manifested Inabilities*																							*Number of Inabilities per Student*
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
1														x						x					2
2					x	x							x							x					4
3	x																			x					2
4	x													x						x					3
5			x																	x					2
6				x	x			x	x		x	x	x	x	x		x		x	x				x	13
7	x										x		x						x	x		x			6
8	x																	x	x	x					4
9					x	x							x		x				x	x					6
10					x	x						x								x					4
11	x				x	x		x			x				x				x	x			x	x	10
12	x					x	x	x						x					x	x					7
13					x						x	x		x	x					x		x			7
14	x												x		x					x					4
15	x								x		x				x			x	x	x	x	x		x	10
16			x	x	x	x					x	x							x	x					8
17	x			x			x	x						x	x	x	x	x	x	x					11
18	x					x		x					x				x			x				x	7
19	x		x		x	x	x								x				x	x					8
20	x				x								x	x	x			x	x	x					8
21		x									x		x	x	x					x			x		7
22	x	x		x		x	x		x				x	x					x	x	x				11
23	x				x								x		x			x	x	x				x	8
24	x			x									x		x				x	x				x	7
25	x		x												x				x	x					5
26	x	x	x		x						x				x			x	x	x					9
27	x		x	x	x	x									x				x	x					8
28		x	x	x							x		x	x				x	x	x				x	10
29	x										x		x	x	x			x	x	x					8
30					x	x	x						x	x				x	x					x	8
31				x									x		x				x	x					5
32												x	x	x				x	x	x				x	7
33	x				x								x		x				x						5
34					x								x					x	x	x				x	6
35					x						x	x			x					x		x			6
36																				x					1
37	x	x	x			x		x			x		x	x	x				x	x				x	12
38	x					x								x	x				x	x				x	7
39			x	x	x		x		x			x								x				x	8
TOTAL	22	5	9	9	17	13	6	6	4		12	7	18	15	22	1	3	11	25	38	2	4	2	13	
PERCENT	56	13	23	23	44	33	15	15	10		31	19	46	38	56	3	8	28	64	97	5	10	5	34	

TESTS OF COMPOSITE TYPE

The Tressler English Minimum Essentials Test purposes to test the pupil's mastery of the essentials in English. The range of the test covers grades 6-12. The time required to give it is

approximately 40-45 minutes. The test consists of 7 tests under the following heads: Test 1, Grammatical Correctness (20 sentences with a choice of 2 expressions); Test 2, Vocabulary (15 sentences with a choice of one synonym from 5); Test 3, Punctuation and Capitalization (5 sentences to be punctuated correctly without rewriting); Test 4, The Sentence and Its Parts (4 sentences in which the subject and verb are to be designated; 3 sentences with words underscored—these to be identified as parts of speech; and 3 sentences to be classified); Test 5, Sentence Sense (10 sentences or groups of words which are to be scored as incomplete, 1, 2, or 3 sentences); Test 6, Inflection and Accent (10 sentences involving the use of different tenses of the verb, possessive singular and possessive plural of nouns, and accent of syllables); Test 7, Spelling (15 blanks for dictated words). Forms A, B, and C are available.

The manual of directions gives the directions for administering and scoring the test, and norms for grades 8-12 by 25 percentile, median, and 75 percentile. No statement is given of its reliability or validity. No suggestions are given for the handling of test scores or the treatment of results in terms of diagnosis and remedial work. Objectives 1, 2, 4, 5, and 6 are partially measured by this test.

Examples of each type of test are given:

Test I. Grammatical Correctness[15]

[The directions given are for the test as a whole, though in the illustrations only parts instead of the whole of certain sections of the test are given.]

Right..... Wrong..... Test I Score.....

In each sentence cross out the incorrect word or expression of the two in parenthesis.

[15] From English Minimum Essentials Test, Form A, devised by J. C. Tressler, published by the Public School Publishing Co.

1. She looks (beautiful, beautifully) in a white dress.
2. The costliness of his armor and apparel (are, is) seen at a glance.
3. The man with eight of his friends (was, were) at the game.
4. They have stiff bristles, which (pick, picks) up the dirt and thread.

Test II. Vocabulary

Test II Score.....

In each sentence look at the underscored word. Then find in the next line a word which means the same or almost the same as the underscored word. Draw a line under this word.

1. I caught a glimpse of the azure sky.
   gray, dark and stormy, clear blue, bright, fiery red
2. He talked frankly with the people.
   roughly, slowly, freely, frequently, convincingly
3. In the trees I saw a cardinal.
   red bird, blue bird, song sparrow, owl, wood pecker
4. Humanity in the mass will probably remain frivolous.
   generous, courageous, silly, cordial, sincere

Test III. Punctuation and Capitalization

Right..... Wrong (other changes).....

Test III Score.....

Punctuate and capitalize these sentences correctly. Do not rewrite the sentences. An unnecessary change is an error. Do not divide one good sentence into two sentences.

1. Although he did not have the courage to stand up and address a small group of his friends he was fearless when he went into battle.

Test IV. The Sentence and Its Parts

Test IV Score.....

Draw one line under the subject word and two lines under the verb in each sentence.

1, 2. In one corner stood two artists discussing the pictures and statues.

Write complex, simple, or compound after each of the next sentences to indicate its class.

8. Scotland possesses a local-option law; and on a request signed by the required number of voters, a borough may vote for or against saloons.

Test V. Sentence Sense

Test V Score.....

The Test

Indicate by 0, 1, 2, or 3 the number of complete sentences in each of the following. You may punctuate and capitalize the exercises if this work will help you to find the number of sentences. What counts, however, is the 0, 1, 2, or 3 which you place at the beginning of each.

( ) 1. We are sending you an exact duplicate of the order blank you sent in so that you may see for yourself whose error caused the delay

( ) 2. Do you know that rowdyism is common among boys who travel to and from school by car they pretend to be gentlemen but lack the essentials of gentility

Test VI. Inflection and Accent

Test VI Score.....

Insert in the blank spaces the verb forms asked for:

1. The dog _____ his leg (past perfect active of *bite*).

Write the possessive singular of this word:

4. it

Place the accent mark over the accented syllable of each of these words:

10. ap pli ca ble (applicable)

Test VII. Spelling

Test VII Score.....

[Words for spelling]

1. lose
2. secretary
3. proceed

An analysis of the results of using this test with high school graduates, with some pointed suggestions as to the significance of these results to high school teachers of English and school administrators, is given by Ashbaugh,[16] who points out types of persisting disabilities and suggests:

Again let me say that one may raise the question as to the propriety of the test. Should one expect high-school seniors to know these language forms well enough to be able to distinguish correct from incorrect form when presented side by side in sentences? Should one expect young people of this age to know the meaning of the words given well enough to select approximate synonyms from groups of words containing synonym, antonym, and miscellaneous words? Should one expect knowledge of subject and verb, of parts of speech, and kind of sentence? Should these students at the end of their high school career be able to recognize a sentence when it is in front of them even though it has neither capital letter at the begin-

[16] "Senior High School English as Revealed by a Standard Test," Bulletin of the School of Education, Indiana University, High School Principals' Conference, Vol. II, No. 2 (November, 1925), pp. 17-18.

ning nor period at the end to separate it from other sentences or groups of words? Should they be able to recognize that a group of words which has no finite verb is not a sentence? Should they be able to give correct verb forms on demand or spell correctly singular and plural possessives? The answer must be sought elsewhere than in these test papers for they say emphatically that one-third to one-half of these students cannot do these things.

Since they cannot do these things, we may well ask why. Is it because they have never been taught? Are these things not in the textbooks and the courses of study? If they are not, then it is no wonder the students cannot respond correctly. But if they are, then the responsibility for this inability lies at the doors of the schools. An examination of these sources finds them present. Our textbook-makers, while giving varying emphasis to these phases of English, all find place for them in their books. Our courses of study show the same status as the textbooks. The inference would seem to be, therefore, that we have not taught these things effectively. I think it is fair to say that we have not considered them of really prime importance. We are beginning to undertake, and to succeed fairly well, to teach the spelling of the most needed words so well that 90 per cent or more of the children will respond correctly whenever called upon. We have placed certain skills in arithmetic and certain facts in various fields of knowledge in this category. We have said that they are so important that we cannot afford to let any but the most inferior of our children leave school without knowing these things, without possessing these skills. These tests so far as this group of students are concerned would show that these abilities in English have not been so regarded. Whether they should be or not is left entirely to your judgment. Perhaps your students would make very high, perfect scores in these lines. Perhaps not. If you believe these things are truly important, it will be very much worth your while to see whether the results are being secured.

The following chart represents the standing of an individual pupil as revealed by Tressler's English Minimum Essentials Test compared to the norms established.

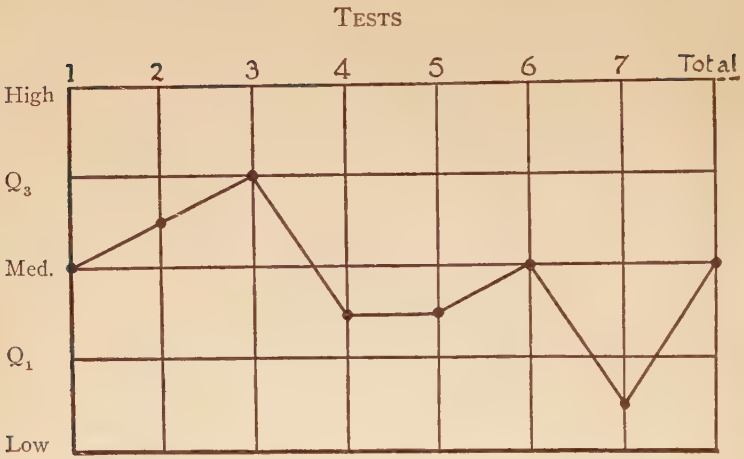


FIG. 10.—The Standing of a Pupil on Various Parts of the Tressler Minimum Essentials Test.

TESTS IN LITERATURE

Tests in the field of English Literature that attempt to measure the expected outcomes of instruction would test (1) ability to appreciate, (2) knowledge of content of the standard literature taught. This knowledge would be in terms of knowledge of the books of best authors, of characters of significance in literature, and of the theme expressed.

The Abbott-Trabue Exercises in Judging Poetry are an attempt to measure the critical judgment of poetry. Series X and Y are available. The manual on "A Measure of Ability to Judge Poetry" is rather complete with, perhaps, a lack in discussion of the use of results. The coefficient of reliability between X and Y series ranges from .037 with seventh and eighth grades to .723 with graduate English students. The test is somewhat short and not satisfactorily scored for the best reliability results. The test consists of 13 sets of four poems

each. One of the four poems is the original one and the other three are alterations of this poem. The sets included are those selected by the judgment of a large number of people somewhat expert in the field of poetry. The pupil is told to "read the poems A, B, C, D, trying to think how they would sound if read aloud. Write "Best" on the dotted line above the one you like best as poetry. Write "Worst" above the one you like least."

Set 3. Sea Shell[17]

A (.....)

Sea Shell, please sing me a song
Of ships and sailor-men;
Of strange kinds of birds and trees
On the Spanish Main:
Of fish and seaweed in the sea,
And whatever creature there may be,—
Sea Shell, please sing me a song!

B (.....)

Sea Shell, Sea Shell,
Sing me a song, O please!
A song of ships, and sailor men,
And parrots, and tropical trees,
Of islands lost in the Spanish Main
Which no man ever may find again,
Of fishes and corals under the waves,
And sea-horses stabled in great green caves.
Sea Shell, Sea Shell
Sing of the things you know so well.

C (.....)

Tender, tender Sea Shell,
Wilt thou sing me, please,
Of thy happy, happy home
'Neath the tropic trees?
Ah, the coral islands!
Ah, the wondrous fish!
For such a song I'd give thee, dear,
Whate'er a Shell could wish.

D (.....)

Sea Shell, I ask you will
You sing a song, please.
All about the ships and sailors
And the parrots in their tropical trees.
The Islands I have read about on the Spanish Main
That no one will see again,
The fish and coral under the wave,
Sea horses that have their stables in caves;
Sea Shell, I ask you will
You sing a song, please.

The correct version is B.

[17] From the Abbott-Trabue Exercises in Judging Poetry, published by the Bureau of Publications, Teachers College, Columbia University.

LANGUAGE TESTS

ENGLISH COMPOSITE

Columbia Research Bureau English Test. By H. R. Steeves, Allan Abbott, and B. D. Wood. Published: 1926. Purpose: to measure achievement in English. Range: high school and college. Time to give: 1 hour, 25 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.40 per package of 25 examination booklets; specimen set, 25 cents.

Cross English Test. By E. A. Cross. Published: 1924. Purpose: to measure pupil's ability to use correct grammatical forms of English and acceptable sentence structure in speaking and writing. Range: high school seniors and college freshmen. Time to give: 40 minutes. Time to score: 4 minutes. Standardized.
Publisher: World Book Co.
Price: $1.20 per package of 25 examination booklets; specimen set, 25 cents.

Iowa Placement Examinations. By George Dinsmore Stoddard and M. F. Carpenter. Published: Form A, 1925; Form B, 1926. Purpose: to measure previous training in certain subjects, and to measure those factors which contribute to subsequent success in these subjects. Range: college freshmen. Time to give: 40 to 45 minutes. Time to score: 2 minutes. Standardized.
Publisher: Extension Division, University of Iowa.
Price: $3.50 per 100 copies of each series complete, plus transportation; specimen copy, either form, 5 cents postpaid.

Purdue English Test. By G. C. Brandenburg. Published: 1919. Purpose: to measure the ability of college freshmen in English. Range: high school and college. Time to give: 35 minutes. Time to score: 15 minutes. Standardized.
Publisher: G. C. Brandenburg, Purdue University.
Price: 2 cents per copy.

Tressler English Minimum Essentials Test. By J. C. Tressler. Published: 1924. Purpose: to test the pupil's mastery of the essentials in English. Range: grades 6-12. Time to give: no time limit; approximately 40 to 45 minutes. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: 75 cents per package of 25; specimen set, 10 cents.

ENGLISH USAGE

Charters' Diagnostic Language Tests. By W. W. Charters. Published: 1918; revised, 1922. Purpose: to test pupils' knowledge of language and to diagnose difficulties. Range: grades 3-12. Time to give: no time limit. Time to score: 2 minutes for each test. Standardized.
Publisher: Public School Publishing Co.
Price: 80 cents per 100, each part; specimen set, 10 cents.

Guiler Diagnostic Test in Punctuation. By Walter S. Guiler. Published: 1926. Purpose: to discover common errors in punctuation. Range: grades 3 through college. Time to give: 10 minutes. Time to score: 3 minutes. Standardized.
Publisher: Rand McNally and Co.
Price: 45 cents per pad of 50 sheets.

Kirby Grammar Test. By Thomas J. Kirby. Published: 1920. Purpose: to measure pupils' ability to choose correct English usage and to recognize the principle in accordance with which the usage is correct. Range: grades 7-12. Time to give: about 35 minutes. Time to score: short. Standardized for grades 7 and 8.
Publisher: Extension Division, University of Iowa.
Price: $1.75 per 100, complete; specimen set, 10 cents postpaid.

Los Angeles Diagnostic Tests: Language. By Caroline Armstrong. Published: 1927. Purpose: to measure achievement in capitalization, punctuation, and language usage. Range: grades 3-9. Time to give: 30 minutes. Time to score: short. Standardized.
Publisher: Research Service Co.
Price: $3 per 100, any form; $25 per 1,000; specimen set, complete, 10 cents.

New York English Survey Tests. By E. L. Cornell, W. W. Coxe, J. S. Orleans, and E. B. Richards. Published: 1925. Purpose: to measure achievement in English. Range: grades 4-8. Time to give: 6 to 21 minutes, according to part given. Standardized.
Publisher: Public School Publishing Co.
Price: $1 per 100, each test; specimen set, 15 cents.

The Schutte English Diction Test. By T. H. Schutte. Published: 1926. Purpose: to test pupil ability to recognize and to point out errors of diction. Range: junior and senior high schools, normal schools, and colleges. Time to give: approximately 60 minutes. Time to score: 3 minutes. Not standardized.
Publisher: Public School Publishing Co.

Price: $3 per package of 25 tests, 25 manuals, and a set of scoring stencils; specimen set, without scoring key, 15 cents.

The Wakefield Diagnostic English Test. By Ethel Wakefield. Published: 1923. Purpose: to measure achievement and diagnose pupil difficulty. Range: grades 9-12. Time to give: 35 minutes. Time to score: 3 minutes. Not standardized.
Publisher: Bureau of Administrative Research, College of Education, University of Cincinnati.
Price: $2 per package of 100 test booklets, complete; in quantities less than 100, 2½ cents per copy; specimen set, both forms, complete, 10 cents, postpaid.

Wilson Language Error Test. By G. M. Wilson. Published: 1923. Purpose: to measure pupil's ability to detect common language errors. Range: grades 3-12. Time to give: 5 to 15 minutes. Time to score: about 2 minutes. Standardized.
Publisher: World Book Co.
Price: $1.25 for package of 25 examination booklets, complete; specimen set, complete, 15 cents, postpaid.

COMPOSITION

Briggs' English Form Test. By Thomas H. Briggs. Published: 1921. Purpose: to measure the pupil's knowledge of the common elements of English composition. Range: grades 7-9. Time to give: no time limit. Time to score: short. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: 80 cents per 100 copies, Alpha or Beta; instructions, 25 cents.

Driggs-Mayhew National Scales for Measuring Compositions. By Howard R. Driggs and Alfred F. Mayhew. Published: 1927. Purpose: to evaluate pupils' compositions. Range: grades 7-9. Time to give: 15 minutes. Time to score: short. Standardized.
Publisher: The University Publishing Co.
Price: 25 cents per copy.

Harvard-Newton Composition Scales. By Frank W. Ballou. Published: 1914. Purpose: to measure the ability of pupils in 4 types of written composition. Range: grades 7-9. Time to give: short. Time to score: short. Standardized.
Publisher: Harvard University Press.
Price: 50 cents per copy of complete scale; separate scales, 12 cents each.

Hudelson's English Composition Scale. By Earl Hudelson. Published: 1921. Purpose: to serve as an objective measure for

grading English compositions. Range: grades 4-12. Time to give: 20 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: 25 cents.

Lewis' Scale for Measuring Special Types of English Composition. By Erwin E. Lewis. Published: 1922. Purpose: to measure ability in letter writing as well as in narrative composition. Range: grades 4-12. Time to give: 20 to 30 minutes. Time to score: 2 to 3 minutes. Standardized.
Publisher: World Book Co.
Price: 25 cents each.

Pressey Diagnostic Tests in English Composition. By S. L. Pressey, Mrs. E. V. Bowers, Helen Ruhlen, and F. R. Conkling. Published: 1923. Purpose: to discover particular weaknesses of pupils in English composition. Range: grades 7 through college. Time to give: 1 to 15 minutes, depending on the test given. Time to score: 1 minute for each test. Standardized.
Publisher: Public School Publishing Co.
Price: Test A, 75 cents per 100; Test B, 75 cents per 100; Test C, $1.50 per 100; Test D, $1.50 per 100.

Seaton-Pressey Diagnostic Tests in English Composition. By J. T. Seaton and S. L. Pressey. Published: 1926. Purpose: to test pupils' ability to perform skills necessary in composition writing. Range: grades 3-6. Time to give: 16 minutes. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: 75 cents per 100 copies of either sheet; specimen set, 10 cents.

Willing Scale for Measuring Written Composition. By M. H. Willing. Published: 1920. Purpose: to measure "story value" and "form value" in composition. Range: grades 4-9. Time to give: 20 minutes. Time to score: 5 minutes. Standardized.
Publisher: Public School Publishing Co.
Price: 9 cents per copy; 3 or more copies, 6 cents each.

LITERATURE

Abbott-Trabue Scales for Judging Poetry. By Allan Abbott and M. R. Trabue. Published: 1921. Purpose: to measure poetry appreciation. Range: grades 7-12. Time to give: 45 minutes. Time to score: no information. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: 5 cents per copy, Series X or Y; Manual of Directions, 25 cents.

Barrett-Ryan Literature Test. By E. R. Barrett and Teresa M. Ryan. Published: no information. Purpose: to test a pupil's knowledge of 35 widely read selections in literature. Range: junior-senior high school and college. Time to give: 35 minutes. Time to score: about 4 minutes. Standardized.
Publisher: Bureau of Educational Measurements and Standards, Kansas State Teachers College.
Price: 75 cents per 25 copies, $1.25 per 50, $2.25 per 100, postpaid; specimen set, 20 cents, postpaid.

Logasa-Wright Seven Tests for the Appreciation of Literature. By Hannah Logasa and Martha M. Wright. Published: 1926. Purpose: to determine the pupil's ability to appreciate literature. Range: third year high school through second year college. Time to give: 12 to 19 minutes, according to the test given. Time to score: no information. Standardized.
Publisher: Public School Publishing Co.
Price: 10 cents per package, containing one set of the seven tests.

Moe Book Tests. By Maurice W. Moe. Published: 1927. Purpose: to test a pupil's comprehension of outside reading books. Range: grades 7 through college. Time to give: no information. Time to score: no information. Standardized.
Publisher: Kenyon Press Publishing Co.
Price: 5 cents per copy; $3.75 per set of tests on 100 commonly read books.

Odell Scales for Rating Pupils' Answers to Nine Types of Thought Questions in English Literature. By Charles W. Odell. Published: 1927. Purpose: to measure achievement in English literature. Range: junior and senior high school. Time to give: no time limit. Time to score: short. Standardized, tentatively.
Publisher: Bureau of Educational Research, University of Illinois.
Price: no information.

Satterfield Objective Tests in English. By Mabel S. Satterfield, Salibelle Royster, Emma Schaedel, Edgar A. Carney, and Wightman F. Melton. Published: 1926. Purpose: to measure comprehension of high school literature books. Range: grades 8-12. Time to give: 20 to 30 minutes. Time to score: 4 minutes. Standardized: no information.
Publisher: Smith, Hammond and Co.
Price: $1.25 per package of 35 tests.

MISCELLANEOUS

Briggs Dictionary Test. By Thomas H. Briggs. Published: 1923. Purpose: to reveal the need for instruction in the uses of a dictionary. Range: junior and senior high school. Time to give: about 2 hours for grade 7 and about 80 minutes for grade 12. Time to score: no information. Not standardized. Publisher: American Book Co.
Price: not yet determined.

Clark Letter-Writing Test. By F. Leslie Clark. Published: 1926. Purpose: to determine a pupil's ability in writing business and social or friendly letters. Range: grades 5-12. Time to give: no time limit. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: $3 per 100 copies; specimen set, 15 cents.

Dolch Study Outline Test. By F. Dean McClusky and Edward W. Dolch. Published: 1926. Purpose: to test pupils' ability to recognize and indicate the outline used by an author in writing a certain topic. Range: high school. Time to give: no time limit. Time to score: 3 minutes. Standardized.
Publisher: Public School Publishing Co.
Price: 75 cents per package of 25 copies of Test I, 15 copies of Test II, and 10 copies of Test III; specimen set, 10 cents.

Gates Graded Word Pronunciation Test. By Arthur I. Gates. Published: 1924. Purpose: to measure ability to pronounce ordinary English words correctly, and to diagnose difficulties. Range: grades 1-6. Time to give: 10 to 20 minutes. Time to score: 1 minute. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: 50 cents per 100, each form.

BIBLIOGRAPHY

Abbott, A. and Trabue, M. R. "A Measure of Ability to Judge Poetry," *Teachers College Bulletin*, Fourteenth Series, No. 2, September 23, 1922, Columbia University.

Ashbaugh, E. J. "The Measurement of Language. What is Measured and its Significance," *Journal of Educational Research*, Vol. 4: pp. 32-39, June, 1921.

Ashbaugh, E. J. "Senior High School English as Revealed by a Standardized Test," *Journal of Educational Research*, Vol. 13: pp. 249-58, April, 1926.

Breed, F. S. and Frostic, F. W. "Scale for Measuring the General

Merit of English Composition in the Sixth Grade," *Elementary School Journal,* Vol. 17: pp. 307-25, January, 1917.

Briggs, T. H. "English Composition Scales in Use," *Teachers College Record,* Vol. 23: pp. 423-52, November, 1922.

Briggs, T. H. "An English Form Test," *Teachers College Record,* Vol. 22: pp. 1-11, January, 1921. *Teachers College Bulletin,* Fourteenth Series, No. 6, December, 1922.

Charters, W. W. "Constructing a Language and Grammar Scale," *Journal of Educational Research,* Vol. 1: pp. 249-57, April, 1920.

Dolch, E. W. "The Measurement of High School English," *Journal of Educational Research,* Vol. 4: pp. 279-86, November, 1921.

Gainsburg, J. C. "Fundamental Issues in Evaluating Composition," *The Pedagogical Seminary,* Vol. 31: pp. 55-77, March, 1924.

Guiler, W. S. "Diagnosing Student Short-comings in English Composition," *Journal of Educational Research,* Vol. 14: pp. 112-19, September, 1926.

Hammond, E. L. "A Study of the Reliability of an Objective Examination in Ninth Grade English," *School Review,* Vol. 35: pp. 45-51, January, 1927.

Hillegas, M. B. "A Scale for the Measurement of Quality in English Composition for Young People," *Teachers College Record,* Vol. 13: pp. 1-54, September, 1912.

Hudelson, Earl. "Diversity of Judgment upon Standards of Content and Achievement in English," *Teachers College Record,* Vol. 27: pp. 33-51, September, 1925.

Hudelson, Earl. "Effect of Objective Standards upon Composition Teachers' Judgment," *Journal of Educational Research,* Vol. 12: pp. 329-40, December, 1925.

Hudelson, Earl. "English Composition: Its Aims, Methods, and Measurement," *The Twenty-second Yearbook of the National Society for the Study of Education,* Part 1, Public School Publishing Company, 1923.

Hudelson, Earl. "Making a Local Composition Scale," *School Review,* Vol. 33: pp. 601-09, October, 1925.

Inglis, A. *Principles of Secondary Education,* Chap. 14, Houghton Mifflin Company, Boston, 1918.

Kirby, T. J. "A Grammar Test," *School and Society,* Vol. 11: pp. 714-19, June 12, 1920.

Leonard, S. A. "Building a Scale of Purely Composition Quality," *English Journal,* Vol. 74: pp. 760-75, December, 1925.

Leonard, S. A. "Wisconsin Tests of Grammatical Correctness," *English Journal,* Vol. 15: pp. 348-57, May, 1926; pp. 430-42, June, 1926.

Lewis, E. E. "Scales for Measuring Special Types of English Composition," reviewed by M. J. Van Wagenen, *Journal of Educational Research,* Vol. 4: p. 427, December, 1921.

McClusky, F. D. and Dolch, E. W. "Study Outline Test," *School Review,* Vol. 32: pp. 757-72, December, 1924.

McGraw, H. W. "Use of Test Data as Basis for Drill in Grammar," *Journal of Educational Research,* Vol. 10: pp. 291-96, November, 1924.

McPhee, Clare. "The Teaching of Language Forms," *Elementary School Journal,* Vol. 26: pp. 137-46, October, 1925.

Morrison, J. C. "The Use of Standard Tests and Scales in the Plattsburg High School," *University of the State of New York Bulletin,* No. 784, July 15, 1923.

Price, Thompson, and Richards. "Translation English," *School and Society,* Vol. 23: pp. 51-56, January 9, 1926.

Ruhlen, H. V. "Experiment in Testing Appreciation," *English Journal,* Vol. 15: pp. 202-09, March, 1926. Reply, Daringer, H. F. Vol. 15: pp. 605-08, October, 1926.

Sackett, L. W. "Comparable Measures of Composition," *School and Society,* Vol. 5: pp. 233-39, February 24, 1917.

Stoddard, George D. "Iowa Placement Examinations," *University of Iowa Studies in Education,* Vol. 3: No. 2, August 15, 1925.

Theison, W. W. "Improving Teachers' Estimates of Composition Specimens With the Aid of the Trabue Nassau County Scale," *School and Society,* Vol. 7: pp. 143-50, February 2, 1918.

Van Wagenen, M. J. "Achievement of College Students in Freshman Rhetoric," *Educational Administration and Supervision,* Vol. 12: pp. 603-17, December, 1926.

Van Wagenen, M. J. "Minnesota English Composition Scales; Their Derivation and Validity," *Educational Administration and Supervision,* Vol. 7: pp. 481-509, December, 1921.

Welch, R. A. "Hillegas Scale as a Measure of Composition Work" (Nassau County Supplement), *English Journal,* Vol. 15: pp. 618-20, October, 1926.

Willing, M. H. "Individual Diagnosis in Written Composition," *Journal of Educational Research,* Vol. 13: p. 77, February, 1926.

Willing, M. H. "Valid Diagnosis in High School Composition," *Teachers College Contributions to Education,* No. 230, Bureau of Publications, Teachers College, Columbia University, 1926.

CHAPTER VIII

Spelling and Handwriting

Spelling is very closely related to language work; however, it is usually treated as a separate subject in the curriculum. The time ordinarily given to it as a formal subject is indicated by the time allotted it in the daily program of the school. Time allotment for spelling in a group of schools is shown by Table 19.[1]

Table 19.—Average Number of Minutes Per Week Allotted to Spelling

Grade	*Minutes per Week*
5	87.5
6	78.2
7	38.9
8	31.3
9	26.0

The objectives usually given for the teaching of spelling in the elementary schools are the following:[2]

1. To develop the ability to spell correctly the words most commonly needed for expression of thought in writing.
2. To develop the meaning and use of words to be spelled.
3. To develop the ability to recognize readily correct and incorrect spelling of words.
4. To develop a desire to spell correctly.

Some objectives in the teaching of spelling are not measured in the standard tests of spelling ability or spelling achieve-

[1] J. M. Glass, *Curriculum Practice of the Junior High School and Grades 5 and 6,* Supplementary Educational Monograph No. 25, p. 107 (adapted).

[2] Adapted from the Indiana State Course of Study, 1926.

ment. The first objective stated—the ability to spell correctly the words most commonly needed for expression of thought in writing—may be rather adequately measured. The second objective—to develop the meaning and use of words to be spelled—is not directly measured by the use of standard tests in spelling. The third—to develop the ability readily to recognize correct and incorrect spelling of words—may be measured by a particular type of test in spelling. This particular method is that of submitting correct and incorrect spelling of the same word and having the pupil decide which is spelled correctly. An example of this method is to list "seperate" and "separate" and ask the pupil to check the correct spelling. Another method is to have the pupil correct any misspelled words found in written work, a method sometimes used as a measure of spelling ability. The fourth objective—to develop a desire to spell correctly—is not measured except indirectly.

Certain difficulties are encountered in attempting the measurement of spelling ability. Should spelling ability be considered as ability to spell a word when one's attention is on the word to be spelled, or when the attention is primarily on the idea to be expressed? Ability to spell correctly when the attention is on the word to be spelled might be called the maximum spelling ability, or the conscious spelling ability; whereas the ability to spell words when attention is on the idea involved might be called the unconscious spelling ability. The latter, of course, is the one by which our success in composition or written expression is determined. This distinction in kind of spelling ability furnishes explanation of why some children who test well on dictated word lists often misspell the same words in composition due to so-called "carelessness." Carelessness in spelling is due very many times to lack of attention to the particular word coupled with only partially complete learning.

Another problem of measurement is that concerning oral and written spelling. In years past schools usually considered spell-

ing an oral subject. Spelling is now considered a subject very closely related to written work and, in that case, any teaching and testing of spelling should be in written form. Many words with which children have difficulty in spelling are words that orally sound alike and whose meaning is differentiated in writing only by correct spelling. Such words are *dying, dyeing,* and *their, there.*

Another problem closely related to the one just stated, which must also be considered in the measurement of spelling, is whether an attempt should be made to measure the extent of spelling ability. Should the attempt be made to measure how well children can spell a certain limited list of most commonly used words or to measure how well they can spell a wider range of words used only occasionally? Generally our measurement has been of the first type.

The examiner may play an important part in attempting to measure the ability to spell. The examiner who carefully pronounces and enunciates the words the children are to spell, will give the maximum opportunity for making good spelling scores. This is especially true where words are given from a list without using each in a sentence. The spelling of "asked" as "ask" or the spelling of "government" as "goverment" may be due to the pronunciation and enunciation of the examiner.

Another problem connected with the use of spelling tests is their disagreement in content. Words found in the Ayres Spelling Scale for use in a certain grade may not be found in other spelling scales for use in that same grade. The disparity of content in spelling texts and tests causes some difficulty in attempting to measure achievement in this line.

Spelling tests are usually taken from spelling lists or scales. Three lists that include a very large number of the words most commonly needed in the expression of thought in writing are the Ayres list, Anderson list, and the list from which the Sixteen Spelling Scales for Secondary Schools were developed. The words included in these first two studies were arrived at

by analysis of several types of letters by children and adults, compositions, and other written material. The words of greatest frequency were then included. The Ayres list finally included 1,000 different words. The Anderson list included 3,105 different words. Lists of words used in the development of the Sixteen Spelling Scales for Secondary Schools combined the lists developed by several investigators, then omitted all of the Ayres list of 1,000 as well as Jones' list of "demons," proper names, hyphenated words, and foreign words not listed in a standard dictionary. A final list was then made of two thousand most frequently used words. From this list the spelling scales were derived.

The Iowa Spelling Scales are made up of 2,979 words taken from the Anderson study. These words were made into standard lists of spelling words, each list being marked with the per cent of difficulty. The relative difficulty of the words was determined by giving the words to pupils in various schools. In this manner the difficulty of each word was determined for each grade.

If a teacher desires to give a test to pupils of the fifth grade, she may select a column of words from the fifth grade scale. The difficulty of each list is standardized. The following words are taken from the Iowa Spelling Scales, Grade V.[3] Forty-one words are in this list, ten of which appear below.

76 per cent

abstract	capital
answered	cheaper
bigger	checked
built	clerk
bulk	cough

These words, as can be seen, are marked 76 per cent. This means that a fifth grade class, taking these words as a spelling test, should average 76 per cent. Certain pupils may make higher and some lower than 76 per cent. However, the average

[3] Published by the Public School Publishing Company.

child should be able to spell 76 per cent of these words correctly. In giving the test, the teacher dictates the words in list form. Each word should be pronounced distinctly. The teacher or examiner should be sure that all pupils understand the word. The score for each pupil is always expressed in terms of per cent. The per cent of correctly spelled words out of the total number of words dictated is the score. If thirty-two of the forty-one words were spelled correctly, the score would be approximately 76 per cent, the average score. If a pupil spelled thirty-seven of these forty-one words correctly, his score would be approximately 90 per cent. The per cent score of any pupil for any grade can always be compared to the norm given for this list of words. The median per cent made by any particular class can be compared with the norm for that grade. For example, if the median score of a fifth grade class is found to be 82 per cent, this, as will be noted, is above the norm given for this grade. The words given in this list are approximately of such difficulty that a sixth grade, taking the test, would make a median score of 85 per cent. This will give some idea of the relative ability of this fifth grade class.

The spelling test applied in the manner just described is merely a measure of ability to spell these words correctly. No measure is here given of ability to spell other words. The composite ability to spell is made up of a large number of abilities to spell specific words. To measure this composite ability in detail would require a measurement of ability to spell many specific words; it would, in fact, require a measurement of the abilities to spell all of the 2,979 words of the Iowa Scale. However, a list selected as the fifth grade list first discussed furnishes a fair index of spelling ability if children have had an opportunity to learn to spell these words. Since ability in spelling very much depends upon experience in the use of the word or instruction in the spelling of the word, the results of any test applied to children will be very much affected by

the other work of the school and by the words found in the spelling book. A measure of achievement based on the teaching of spelling, both direct and incidental, should be on those words the child has had a chance to learn. Textbooks in spelling do not include the same words in the same grades. This non-uniformity makes it difficult to measure achievement except on the words common to the books mostly used.

The measurement of spelling, as has been described here, does not take into account how difficult a word is to learn, but only how often it is misspelled. However, the words most frequently misspelled are the most important as a basis of remedial work. From the point of view of developing better methods, it would be desirable to have greater knowledge of the relative difficulty of the learning of various words.

For the purpose of measuring the ability of a class it is sometimes advisable to select from various lists words ranging from those that are easy to those that are very difficult. By the use of these words in sentences illustrating their meaning in context, a test of the sentence spelling type can be developed. Monroe developed such a scale by selecting appropriate words from the Ayres list and imbedding them in sentences in which the other words presented no spelling difficulty. He added a timing element in this test based on the rate of handwriting. An example of this test follows.

Test I[4]

Seconds	
60	He *bought* a *railroad ticket* to the city.
60	*Collect* the *account* before Sunday.
55	*Those children* will *return* soon.
46	*Anyway* she is *ready* to go.
24	*Please omit both* names.
60	*Few change* trains here.

[4] From "Timed Sentence Spelling Tests," arranged by W. S. Monroe; words selected from Ayres' list, Column M, for Grades 3 and 4. Published by the Public School Publishing Company.

Seconds

36 He *says* the *great office* is full.
25 *Who died* this morning?
59 The *money* for the *picture* was *paid* to us.

Starch developed tests for measurement of the extent of the pupil's spelling ability.[5] He sampled words from the total vocabulary by taking every first defined word on every even numbered page of the Webster's International Dictionary. From these words tests were developed. These tests are not a measure of the direct results of teaching but rather a test of the range of ability to spell all words of the English vocabulary.

Some words are misspelled very frequently. The 100 most frequently misspelled words of very common usage were listed by Jones in his investigation[6] of spelling. Many of these words are found in other lists. English teachers as well as teachers of spelling often make it a minimum requirement that children shall learn to spell these words. Following are typical words from Jones' "Demons."

any	ache
much	read
beginning	said
blue	hoarse
though	shoes
coming	tonight
early	wrote
instead	enough
easy	truly
through	sugar
every	straight

A school superintendent recently asked his daughter, who was in the sixth grade, to spell all these words for him. The

[5] Daniel Starch, "The Measurement of Efficiency in Spelling, and the Overlapping of Grades in Combined Measurements of Reading, Writing, and Spelling," in *Journal of Educational Psychology*, Vol. 6 (March, 1915).

[6] *Concrete Investigations of the Material of English Spelling*, University of South Dakota Bulletin, 1913.

daughter wrote the words as they were dictated. The father found that every one was spelled correctly. Upon further investigation he found, however, that many of the words that had been taught in the grades below the sixth his daughter could not spell. This incident illustrates two things. One is that, while there may be many spelling demons that are general with almost every one, a particular person may not have to bother with those "demons" and yet have individual "demons" that are as real and bothersome as those listed. The second is that the publishing of a list of so-called "demons" or deficiencies calls attention to a needed emphasis that may be taken up by teachers who both consciously and unconsciously teach better on those points than on others. This emphasis may or may not be profitable, depending upon the specific difficulties of the particular class. Sometimes it is more important for the classroom teacher to know the special difficulties of individuals and of the class than to compare their achievement with a standard. The teacher needs to know what words are most often missed by the class as a whole and by certain individuals. When this information is at hand, the next question is, "How were the words misspelled?"

Table 20 is taken from a study of the results of an application of the spelling test of the Indiana Composite Achievement Test to an eighth grade class.

From the results of this test as shown in the table, the following points may be noted: (1) that certain words are not misspelled by these pupils to any marked degree; (2) that certain other words are misspelled by a rather large per cent of these pupils; (3) that certain pupils are very deficient in the ability to spell most of these words. The words *journal, president,* and *important* offer little difficulty. The words *interfere, lovable,* and *crochet* are misspelled by a relatively large per cent of the pupils. Pupils G.R. and J.H. missed a rather large per cent of these words, while pupil M.M. did not misspell any of them.

TABLE 20.—DIAGNOSTIC CHART OF SPELLING ACHIEVEMENT OF 8TH GRADE PUPILS, INDIANA COMPOSITE ACHIEVEMENT TEST, MAY, 1927

Pupil	26. lovable	27. schedule	28. important	29. insurance	30. employ	31. agreement	32. announce	33. familiar	34. principal	35. recognize	36. universal	37. interfere	38. convenient	39. journal	40. ambitious	41. crochet	42. management	43. government	44. absolutely	45. population	46. president	47. recommendation	48. criticize	49. difficult	50. accomplish
B.A.	x							x								x			x						
M.B.	x				x						x					x									x
J.C.	x								x															x	
E.D.	x				x		x	x								x	x								
J.D.	x	x	x			x					x	x				x						x			
M.D.	x									x		x			x	x									
E.H.		x							x	x		x	x		x	x			x						
J.H.	x			x			x	x			x	x	x		x	x		x	x			x		x	
K.H.	x				x					x		x				x			x			x	x		
M.H.	x						x	x			x								x			x			
H.J.	x						x						x			x	x							x	x

A.K.			x	x	x	x		x				x			x	x									
C.M.		x					x					x			x	x									
E.M.	x			x						x		x				x							x		
G.M.		x										x				x									
K.M.								x							x							x			
M.M.																									
F.R.	x						x								x										
G.R.	x	x				x		x	x	x	x	x	x			x	x	x	x	x			x	x	x
M.R.													x									x			
P.R.	x			x					x							x			x			x			
H.S.				x				x								x									
W.S.									x							x			x			x			
M.T.					x		x	x	x							x									
D.W.	x			x			x		x													x			x
TOTAL WORDS MISSPELLED	15	5	1	7	5	3	8	9	7	5	5	11	5	0	7	18	3	2	8	1	0	10	4	4	5
PERCENTAGE OF ERRORS	60	20	4	28	20	12	32	36	28	20	20	44	20	0	28	72	12	8	32	4	0	40	16	16	20

Note—X means word misspelled.

More important, however, than analysis of the class results is a study of the causes of errors made in spelling those words that cause the most difficulty. A study was made of the ways in which the three most misspelled words were spelled. This was done to determine the method of attack needed to remove the difficulty. Following is an analysis of these misspellings:

26	37	41
lovable	*interfere*	*crochet*
loveable (14)	interfier (1)	crochey (4)
lovebul (1)	interfer (7)	crouque (1)
	interfeer (1)	chrosheigh (1)
	interferior (1)	croshey (2)
	interfear (1)	crosque (1)
		crocha (1)
		crosha (1)
		crouquet (1)
		croquet (3)
		croshay (1)
		croquette (1)

A glance at the misspellings of these three words will indicate the specific difficulty that each word offers. In the word *lovable* the misspelling was decidedly of one kind; the letter *e* on *love* was not dropped by the pupils in spelling *lovable*. In the word *interfere* the types of misspellings were more varied than in the case of *lovable;* however, failure to add the final *e* accounted for more than one-half of the mistakes made. The portion of the word "interf" is spelled correctly by all of the pupils; the trouble was in the ending of the word. In the last word, the types of misspellings are much more numerous and varied. A study of the misspellings of *crochet* will, however, suggest two causes. The spelling is like the pronunciation of the word. In general, of course, this is a safe guide to the spelling of a word, but in the case of this word this guide is misleading. In some cases the pupils had evidently learned

that it is not spelled as it is pronounced, but they could not recall the correct spelling.

CAUSES OF MISSPELLING OF WORDS OF COMMON USAGE

Carelessness is a common cause of misspelling. This particular cause should be of great interest to teachers and supervisors. It is often given as a cause and then excused by saying, "Mary knows better." It seems to represent, as has been said before, the condition of inattention and imperfect learning. Pupils sometimes write their compositions and then go back over them and check the spelling of the words. They have learned that they best express the thought by going on with the story to the end. At the same time, these pupils realize that, when their attention is off the spelling of words in the composition, mistakes are likely to happen. Such a type of misspelling is most easily remedied by giving attention to the spelling of the word and then following with attentive drill.

Imperfect audition is the cause of the misspelling of many words. The largest factor contributing to imperfect hearing is poor and improper pronunciation and enunciation on the part of parents, teachers, and children. Imperfect audition may also be caused by physical defects of the auditory organs. So very often misspelled words are spelled as they are heard or pronounced. The most common misspelling of *representative* is *repersenative* or *repersentive*. Observation of the pronunciation of this word will easily give the clew to these spellings. The word is usually pronounced "reper" and not "repre." The middle part of the word is usually omitted or slurred; that is, it is pronounced "senative" and not, as it should be, "sentative." An analysis of more than five hundred misspellings of this word showed practically no one misspelling the "rep" or "tive" part of the word. This type of illustration would apply to many words commonly misspelled. It is an interesting thing to note that mispronunciation by the child

is even more likely to cause this defect than such mispronunciation by the teacher or by the person giving an examination. When the problem of spelling the word is in the mind of the pupil, he usually pronounces the word to himself. This makes for misspelling. Children with defective speech have much difficulty with spelling.

Imperfect vision causes much difficulty in spelling. Physical conditions that cause blurred letters in words cause defective spelling. The length of the word and the similarity of letters in the word are also causes of difficulty. The middle of a long word like *superintendent* is more often misspelled than are the first and last parts of the word. This may be due to faulty vision or to inaccurate perception.

A fourth type of error in spelling might be called faulty association. This is often given as a reason for doing away with both phonetic spelling and spelling by rules. Even when these methods are eliminated, however, pupils still make incorrect associations. A college freshman recently spelled the words *all right* in the following ways: *allwright, allwrite, alright, alwright.* The case was talked over with this below-the-average student. Although it seems rather unusual, he did not know the real difference between the words *right, write,* and *wright;* thus he associated *all* with any one of the three that might come to his mind. Other causes of misspelling inherent in the word are the leaving out of silent letters in a word and the confusion of *ei* and *ie, sc* and *ce, al* and *le.* These errors offer difficulties hard to account for in any way except through the peculiarities of the language.

The following tested remedial devices are of use to any teacher in overcoming difficulties of special cases or the special difficulties of a class.

1. Present the word in as many ways as possible. Have the child hear it, see it, pronounce it, spell it orally, write it, or make any combinations of these methods. Correct presentation by as many sensory modes as possible is good pedagogy of

long standing. Such presentation will require that some words be presented in syllables.

2. Give special attention to the particularly difficult phases of the word. It is doubtful whether this should be done before testing but, as a remedial device, the practice is sound. It brings the child's attention specifically to the difficult letter or part of the word. This is the pupil's particular error. It is the job of the teacher to find the cause of that error and then to help the pupil overcome it. The word *definite* was misspelled by a group of children in a class observed. When the types of misspelling were analyzed, the difficulty was found to be in the letter third from the end. Some pupils had left out the letter, others had used an *a* instead of an *i*. In the writing of this word on the board for the group, the teacher used a red *i* every time she wrote it. The children underlined the *i* on their papers, they emphasized the *i* in pronouncing it. In the remedial teaching, the teacher did not emphasize the error or misspelling, but centered attention on the possible methods of developing correct habits of spelling. Remedial teaching and drill were given on the specific need. In drilling on the right form, the teacher would each time, however, have the whole word rewritten, in order to keep the letter *i* in its proper relationship to the rest of the word. Special attention should always be given to the order of letters in the word.

Words of similar spelling can well be studied together. A particular word difficulty can sometimes be overcome by having the child see that word with words similar in spelling to it, but words that he already knows how to spell. Bringing to the attention of the child the common elements of different words may be helpful.

Teaching meaning and use of words along with proper spelling may not increase the spelling achievement of children so very much; however, its value in attaining the other objectives of the teaching of spelling is so important that it should not be omitted.

HANDWRITING

Handwriting is very closely associated with language communication. Achievement in handwriting both in speed and legibility is an aid in written communication. Writing receives a relatively small amount of time in the daily program of the schools, even though it is much used in school work.

The one big objective of handwriting is "to develop sufficient skill to enable pupils to write easily, legibly, and rapidly enough to meet present needs and social requirements." To this statement might be added "and in such a way as will help develop the art of written communication and desirable social habits of writing."

Black and Russell [7] made a study of the rate and quality of handwriting of 275 pupils who made normal progress from grade to grade for seven consecutive years. The result of this study is shown in Figure 11. This diagram shows the increase in rate and in quality by successive years and grades. The two lines of the chart should not be compared with each other. A point on one line can, however, be compared with another point on the same line.

The noticeable phases of this chart are (1) the rather consistent gain in the rate of handwriting with both age and grade from year to year, (2) the rapid gain in legibility up to and including 5A grade and the apparent slump in gain in legibility after that point. This is not in perfect accordance with the general findings by those who test all pupils in all grades at the same time. In the case of those studies, however, different children represented the various grades.

The emphasis in handwriting has changed from "pretty" writing, with its shaded lines and excessive "style" found mostly in copybooks, to legible writing that is rapid and that

[7] "A Study of Handwriting in Forty Indiana Cities," Twelfth Conference on Educational Measurements (1925), pp. 43-57.

can be produced more easily. By legibility is meant the ease with which the writing can be read. If one reads the hand-

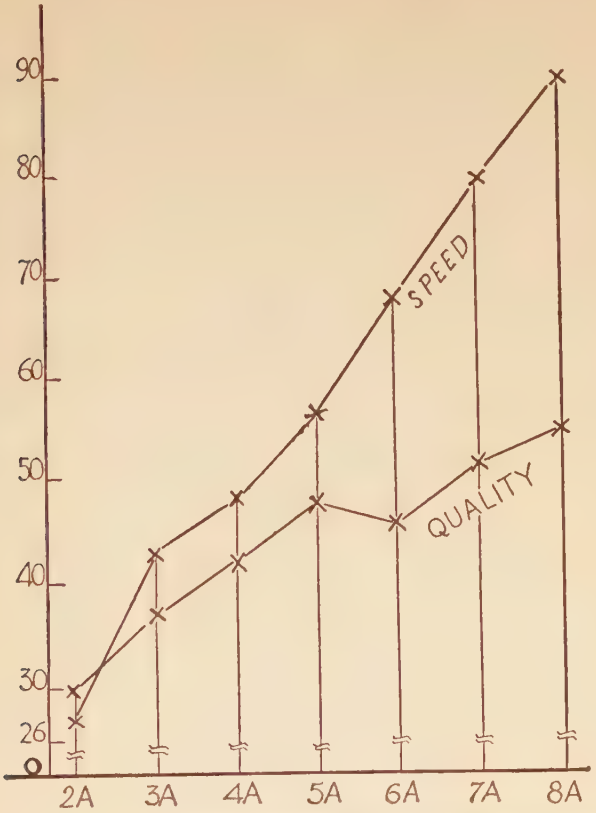


FIG. 11.—Chart Showing Rate and Quality of Handwriting.[8]

writing easily, whether it is backhand or slant in style, it is rated high in legibility. The rate at which the pupil writes, like the rate at which he reads, is important, provided he

[8] *Ibid.*

writes legibly. Too much emphasis on speed is likely to decrease legibility.

The measurement of handwriting for a general survey purpose may be obtained by use of the Ayres Handwriting Scale (Gettysburg Edition).

The Ayres scale is made up of eight samples of children's handwriting that increase in quality as determined by legibility. Samples were obtained by having children write the first part of Lincoln's Gettysburg Address. From the samples of children's handwriting thus obtained competent judges selected eight samples to represent different degrees of quality. The scale made from these selected samples is numbered 20—30—40—50—60—70—80—90 on quality. To use this scale as a test of handwriting, the examiner first teaches the children the first three lines of the Gettysburg Address so they are familiar with it. At a given signal they begin to write with the copy still before them on the board. The children write for exactly two minutes. The writing samples are then scored for rate and for quality. The rate is indicated by the number of letters written per minute. This score is obtained by dividing the total number of letters written by two. As the copy of the Gettysburg address with a letter count is given on the scale, it is easily scored for speed. To score for quality the scorer should "slide each specimen along the scale until a writing of the same quality is found. The number at the top of the scale above this shows the value of the writing to be measured."

Samples of two values of the scale are shown in Figure 12.

The scoring of the pupil's handwriting with this scale will give results of four kinds. The pupils can learn to measure their own work objectively by comparing their own writing to the scale. The teacher can obtain a measure of the standard of the class as compared to the norm. The use of the scale can help to standardize the work in handwriting. The use of the scale will help to select more carefully those who need special diagnosis and help.

FIG. 12.—Samples of Handwriting, with Quality Score. (Published by the Russell Sage Foundation.)

In the chart (Figure 13) four pupils, *a, b, c, d,* who are in the fifth grade, are shown in relation to the standards for the various grades. Pupils *a, b,* and *d* are above or near the standard for their grade. That is, they are near or above the lines that represent the fifth grade standard in rate and legibility. Pupil *c* is barely above third grade standard in both rate and quality.

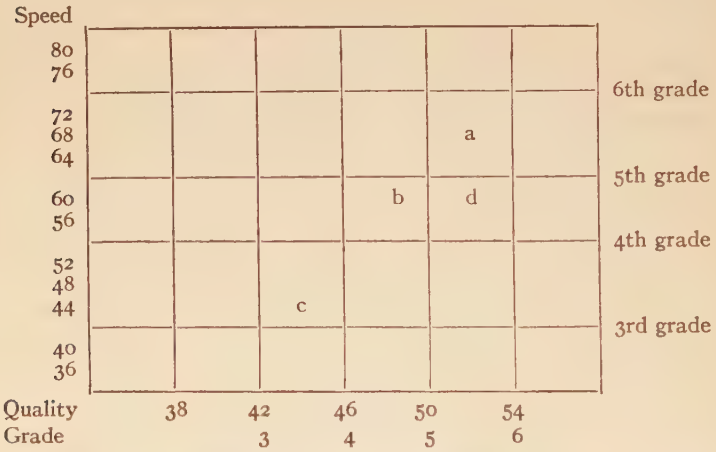
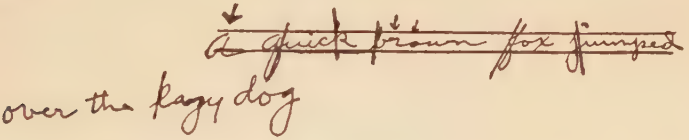


FIG. 13.—The Placement of Pupils on Diagnostic Chart of Ayres Handwriting Scale.

What is the trouble with pupil *c* that he is far below the norm in both rate and quality? By an analysis of his writing, using the Freeman Chart for Diagnosing Faults in Handwriting, the teacher can find the specific difficulties. This rather large chart devised by Freeman designates certain specific qualities of handwriting. The qualities considered are (a) uniformity of slant, (b) uniformity of alignment, (c) quality of line, (d) letter formation, (e) spacing. By the use of this chart particular difficulties can be ascertained. By drawing

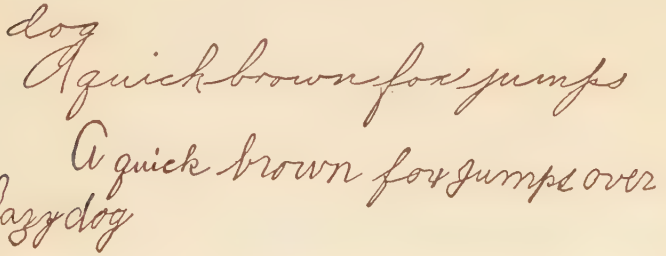
straight lines through the longer letters showing their slant it is soon easy to see how nearly the lines of slant are parallel. Alignment is measured by the degree to which the letters touch two parallel lines drawn at the bottom of the word and at the height of the smaller letters. The qualities of perfect letter formation are many. Are the *n's* and *u's* written properly, are the *o's* and *a's* alike or properly distinguished, are the *o's* closed? These are types of questions that bring out this point.

Following is a sample of handwriting in which the faults in alignment, uniformity of slant, and letter formation are pointed out.



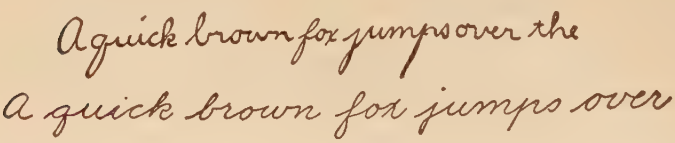
In this example it can be noted that the lines of slant are not the same for various parts of the handwriting. The sample does not fit to either of the lines in the measure of alignment. The faults of letter formation in the sample are very many. Arrows point to outstanding faults.

Here are two samples of quality of line.[9]

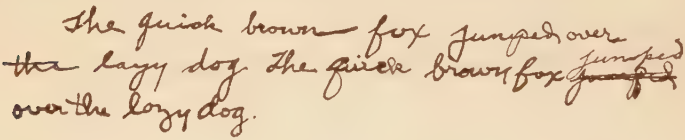


[9] F. N. Freeman, "Freeman Chart for Diagnosing Faults in Handwriting" (Houghton Mifflin Company).

Two samples of spacing are also given.



The causes of poor alignment, slant, and letter formation may be the lack of any fixed habit of writing or a wrong habit firmly fixed. The changeable type of writing is caused by no definite habit of making a letter. A sample is given.



These two sentences are samples of the same person's writing. There are few fixed habits indicated in the examples. An *r* is slurred in one word and consequently has only one point; whereas in another word it is made with two. The *n* in brown is in one case like *u*. The first thing to be done in this case is to put before the child a clear image of the formation of the letter and to have him see the characteristics of the letters and the distinction between them. The child's imitation of correctly formed letters with their distinct characteristics followed by intensive drill on this corrected form, both by single letters and then in words and sentences, is a valuable means of remedial work. In those cases where distinctly bad habits are already formed, an additional difficulty of breaking a habit is involved and must be dealt with.

Practically it may be better to accept defeat in our attempts to develop legible handwriting if such legibility is not attained

by the high school period. It may be much more practical to require that such pupils as are found in high school with legibility below 40 on the Ayres Scale be required to learn to use a typewriter.

Extreme slants in handwriting do much to produce poor legibility. The position of the writer in reference to his paper and to his desk is likely to be a cause of illegibility. Freeman[10] suggests a correct position as follows:

In the first place, the arm should rest with nearly the full length of the forearm upon the desk, with a possible exception to be noted presently. This gives the arm firm support upon the muscle pad on the lower side of the forearm, and it is upon this muscle pad as a sort of rolling base that much of the movement of the arm is executed. If the forearm projects more than three or four inches over the edge of the desk, the weight of the arm is divided between the hand and the shoulder, and the arm, in executing the arm movements, must swing from the shoulder.

Coördination of the muscles and nerves in handwriting is basic to a good product. Cramped writing may be due to a lack of this organization and coördination of movement. This means that freer arm movement, and elimination of much of the finger movement, and a better organization of all phases of movement will tend to give better results. This elimination of excessive finger movement will tend to reduce fatigue as well as a cramped type of handwriting. This does not mean, as is often stated, that exclusive arm movement should be taught, but that coördination and organization of all the movements of finger, hand, wrist, and arm should be the aim of teaching. In remedial teaching it may be necessary to exaggerate one phase more than another to develop its use. A child may use finger movement to the exclusion of arm movement. In this case special emphasis on arm movement is desirable although coordination with other movements should follow.

[10] *The Teaching of Handwriting* (Houghton Mifflin Company), p. 62.

Almost every teacher of handwriting is acquainted with the better methods of teaching handwriting. These are applicable to remedial teaching with this variation. Remedial teaching follows diagnosis that has revealed specific defects; thus the teaching is not on all phases of development but on the points of defect.

TESTS IN SPELLING AND HANDWRITING

SPELLING

Ayres' Spelling Scale. By L. P. Ayres. Published: 1915. Purpose: to measure the ability to spell commonly used words. Range: grades 3-8. Time to give: 4 minutes. Time to score: 20 seconds. Standardized.
Publisher: Russell Sage Foundation.
Price: $9 per 100 copies; specimen copy, 10 cents.

Iowa Spelling Scale. By E. J. Ashbaugh. Published: 1919. Purpose: to measure spelling ability of elementary school pupils. Range: grades 2-8. Time to give: no time limit. Time to score: varies. Standardized.
Publisher: Public School Publishing Co.
Price: $1.25.

Monroe's Timed Sentence Spelling Test. By W. S. Monroe. Published: 1918. Purpose: to measure speed and accuracy in spelling. Range: grades 2-8. Time to give: 6 to 15 minutes according to test given. Time to score: no information. Standardized.
Publisher: Public School Publishing Co.
Price: single sets, 12 cents; 3 or more sets, 10 cents each; extra record sheets, 1 cent each.

Morrison-McCall Spelling Scale. By J. Cayce Morrison and William A. McCall. Published: 1923. Purpose: to measure and compare spelling abilities. Range: grades 2-8. Time to give: about 10 minutes. Time to score: about 1 minute. Standardized.
Publisher: World Book Co.
Price: 25 cents per copy.

Sixteen Spelling Scales. By Earl Hudelson, F. L. Stetson, and Ella Woodyard (under the direction of Thomas H. Briggs and Truman L. Kelley). Published: 1920. Purpose: to measure

spelling ability. Range: grades 7-12. Time to give: about 15 minutes. Time to score: about 1 minute. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: 40 cents each.

HANDWRITING

Ayres' Handwriting Scale (Gettysburg Edition). By L. P. Ayres. Published: 1917. Purpose: to measure speed and quality in handwriting. Range: elementary or high schools. Time to give: 2 minutes. Time to score: 10 seconds. Standardized.
Publisher: Public School Publishing Co.
Price: 18 cents per copy; 3 copies or more, 14 cents each; extra record sheets, 1 cent each.

Freeman Chart for Diagnosing Faults in Handwriting. By F. N. Freeman. Published: 1914. Purpose: to measure slant, alignment, quality of line, letter formation, and spacing. Range: all grades. Time to give: no information. Time to score: no information. Standardized.
Publisher: Houghton Mifflin Co.
Price: 40 cents per copy of scale and complete accessories, postpaid; 3 or more copies, 37 cents each.

Pressey Chart for Diagnosis of Illegibilities in Handwriting. By S. L. and L. C. Pressey. Published: 1927. Purpose: to measure "readability" of handwriting and to diagnose difficulties in writing legibly. Range: grades 3 through college. Time to give: no time limit. Time to score: no information. Standardized.
Publisher: Public School Publishing Co.
Price: 12 cents.

BIBLIOGRAPHY

Alltucker, Margaret M. "Improving Spelling Through Research," *Journal of National Education Association*, Vol. 15: pp. 271-72, December, 1926.

Ayres, L. P. "A Scale for Measuring the Quality of Handwriting of Adults," *Russell Sage Foundation pamphlets*, 1915.

Ayres, L. P. "A Measuring Scale for Ability in Spelling," *Division of Education*, Russell Sage Foundation, 1915.

Breed, F. S. and Culp, V. "Application and Critique of the Ayres Handwriting Scale," *School and Society*, Vol. 2: pp. 639-47, October 30, 1915.

Breed, F. S. "Comparative Accuracy of the Ayres Handwriting Scale," Gettysburg Edition, *Elementary School Journal,* Vol. 18: pp. 458-63, February, 1918.

Breed, F. S. and Down, E. F. "Measuring and Standardizing Handwriting in a School System," *Elementary School Journal,* Vol. 17: pp. 470-84, March, 1917.

Davis, Georgia. "Remedial Work in Spelling," *Elementary School Journal,* Vol. 27: pp. 615-26, April, 1927.

Fourth Yearbook of the Department of Superintendence of the National Education Association, 1926.

Freeman, Frank N. "An Analytical Scale for the Judging of Handwriting," *Elementary School Journal,* Vol. 15: p. 432, April, 1915.

Freeman, Frank N. "Some Practical Studies of Handwriting," *Elementary School Teacher,* Vol. 14: pp. 167-79, December, 1913.

Gates, A. I. *The Psychology of Reading and Spelling,* Teachers College Contributions to Education, No. 129, 1922.

Gates, A. I. and Chase, E. H. "Methods and Theories of Learning to Spell, Tested by Studies of Deaf Children," *Journal of Educational Psychology,* Vol. 17: pp. 289-300, May, 1926.

Gray, C. T. "A Score Card for the Measurement of Handwriting," *University of Texas Bulletin,* No. 37, July, 1915.

Hollingsworth, Leta S. *The Psychology of Special Disability in Spelling,* Teachers College, Columbia University, Contributions to Education, No. 88, 1918.

Hudelson, Stetson, and Woodyard. "Sixteen Spelling Scales," *Teachers College Record,* Vol. 21: pp. 337-92, September, 1920.

Kallom, Arthur W. "Some Causes of Mis-Spellings," *Journal of Educational Psychology,* Vol. 8: pp. 391-406, September, 1917.

Kingsley, John H. "The Test-Study Method Versus the Study-Test Method in Spelling," *Elementary School Journal,* Vol. 24: pp. 126-29, October, 1923.

Kirk, John G. "Handwriting Survey to Determine Grade Standards," *Journal of Educational Research,* Vol. 13: pp. 181-88, March, 1926.

Koos, L. V. "The Determination of Ultimate Standards of Quality in Handwriting for the Public Schools," *Elementary School Journal,* Vol. 18: pp. 423-46, February, 1918.

Lester, John A. "Spelling Ability and Meaning Vocabulary as Indications of Other Abilities," *Journal of Educational Psychology,* Vol. 16: pp. 175-81, March, 1925.

McCoy, J. P. "Year to Year Difference in Spelling Ability Measured in Performance Group Grades," *School and Society,* Vol. 22: pp. 22-23, July, 1925.

Morton, R. L. "The Reliability of Measurements in Spelling," *Journal of Educational Method,* Vol. 3: pp. 321-28, April, 1924.

Orleans, J. S. "Measuring Ability to Spell," *School and Society,* Vol. 23: pp. 407-08, March, 1926.

Rugg, H. O. "Rugg's Scale for Measuring Free-Hand Lettering for Use in Secondary Schools and Colleges," *Journal of Educational Psychology,* Vol. 6: pp. 25-42, January, 1915.

"Sixteen Spelling Scales Standardized in Sentences for Secondary Schools," *Teachers College Bulletin,* Twelfth Series No. 19, May 21, 1921, Columbia University.

Third Yearbook of the Department of Superintendence of the National Education Association, 1925.

West, Paul V. "Improving Handwriting Through Diagnosis and Remedial Treatment," *Journal of Educational Research,* Vol. 14: pp. 187-98, October, 1926.

Whitty, Paul A. "Diagnosis and Remedial Treatment of Poor Spellers," *Journal of Educational Research,* Vol. 13: pp. 39-44, January, 1926.

Woody, Clifford. "The Evaluation of Two Methods of Teaching Spelling," *Studies in Education,* Yearbook No. XV of the National Society of College Teachers of Education, The University of Chicago Press, 1926.

CHAPTER IX

Reading

An outstanding development of the last half century has been the wider use of reading in the lives of pupils and adults in school and out of school. This and other reasons give reading a place of unusual importance in the school program. A glance at the following table (Table 21) will give some idea of the place of reading as a special subject in the school's daily program.

Table 21.—Average Time in Minutes Per Day and Percentage of Time Devoted to Reading in 49 Cities [1]

Grade	*Average Number of Minutes per Grade per Day*	*Percentage of Total Time in Elementary-School Curriculum*
I	84.2	32.6
II	80.8	28.6
III	66.4	22.2
IV	49.0	16.0
V	36.4	11.8
VI	28.2	9.2
VII	28.4	9.0
VIII	27.2	8.7

This table should be read as follows: "In Grade I 84.2 minutes per day are given to reading, and this represents 32.6 per cent of the total time in that grade.

This table, of course, does not indicate the time given to reading in connection with other subjects in the curriculum. The time spent in reading in other subjects combined with the

[1] F. C. Ayer, Bulletin No. 1, Dept. of Research, Seattle Public Schools, 1924.

time spent on the special subject of reading makes up a very large per cent of the total school time.

The major objectives of the teaching of reading are usually stated as follows:[2]

1. To give a rich and varied experience through reading.
2. To develop strong motives for, and permanent interests in, reading.
3. To develop desirable attitudes and economical and effective habits and skills.
   a. To develop well established fundamental reading habits.
   b. To develop effective habits of intelligent interpretation.
   c. To develop effective oral reading.
   d. To develop the ability to use books, libraries, and other sources of information economically and effectively.

The statements in a good course of study set forth these major objectives and also break them into minor grade objectives. The best activities for the attainments by grades are also given in a course of study. It is not possible to give all of these in detail here. Any testing program, however, must take into account these statements of grade objectives, minimum essentials, and pupil activities.

THE READING PROGRAM

The kindergarten and nursery schools attempt to build up experiences of a number and kind that will best prepare the child for instruction in reading. This work may have to be done at the beginning of the first grade if it has not been done in the pre-first-grade life of the child.

The first period of instruction usually given in the first grade has as its purpose "to stimulate a keen interest in reading activities, to cultivate a thoughtful reading attitude, and to develop reasonable speed and accuracy in reading simple pas-

[2] Adapted from Gist, "Elementary School Supervision," Chas. Scribner's Sons; and Anderson and Davidson, "Reading Objectives," Laurel Book Company.

sages."[3] The association of symbol with meaning and sound of words alone and with other words, the articulation of the eye with the mechanics of the book, and then with the page and line, and the securing of some new experiences characterize the development of this grade or period.

The second and third grades usually represent a period of rapid growth in the fundamental skills, habits, and attitudes of reading. This is the period between the time when the child has been taught to read and the time when he must extend his silent reading to content and enriching reading. The child should be able to read independently, to read more rapidly silently than orally, and to interpret the reading material of his grade.

The reading of the fourth, fifth, and sixth grades should continue the growth of those habits developed in grades two and three and provide for a wide and enriching type of experience through such reading as content subjects, travel, biography, and literature.

The reading of the junior and senior high school should extend the experience possibilities and intellectual development of its pupils. The wholesome use of leisure time on one side and the work type of reading on the other call for many different types of reading and reasons for reading. The vocabulary of reading may become specialized and highly developed in either line.

In general the composite ability of reading increases by grades and ages. This fact is shown by a study of norms by grades and ages.

The skills, habits, and abilities involved in reading are many and varied. The psychologist describes the processes of reading as follows: (1) visual impression on the retina, (2) nerve impulses to the brain, (3) associations which give meaning. These are always present in silent reading. If oral reading or

[3] *Twenty-fourth Yearbook* of the National Society for the Study of Education, Part I, p. 35 (1925).

inner speech is carried on, then (4) the motor impulses are aroused, (5) these impulses pass to the muscles of the proper organs causing speech, (6) the words are uttered by these speech organs. There are, however, a very large number of factors that affect this process and determine reading efficiency.

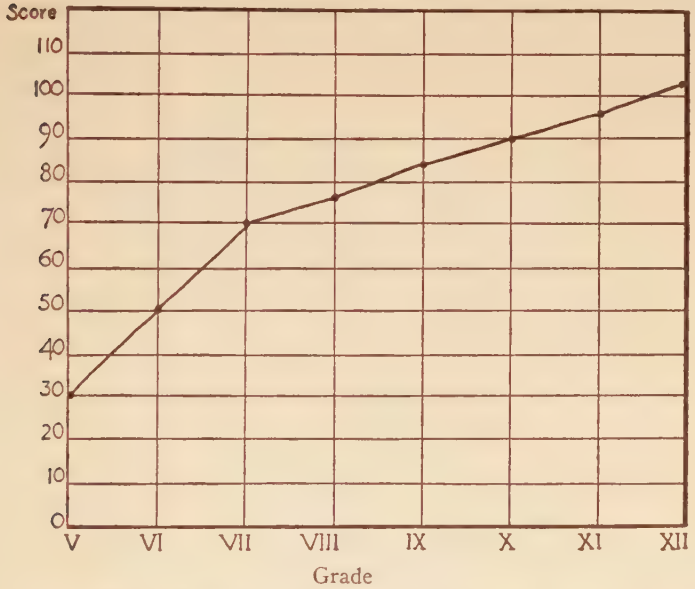


FIG. 14.—The Increase in Scores by Grades on Haggerty Silent Reading Test Sigma III. (Taken from norms.)

The following list by Brooks [4] includes a very large number of these conditions that should be considered in the measurement of reading and in diagnosis.

A. External Factors
   1. Uniformity of length of lines.
   2. Size and kind of type.

[4] *The Applied Psychology of Reading* (D. Appleton and Company, New York, 1926), pp. 91-92.

3. Spacing between words and lines.
4. Kind and color of paper.
5. Illumination.
6. Sentence structure and style.
7. Demands of material upon memory.
8. Vocabulary.
9. Appeal to interest and arousal of effort resulting therefrom. (This is both an internal and external factor.)

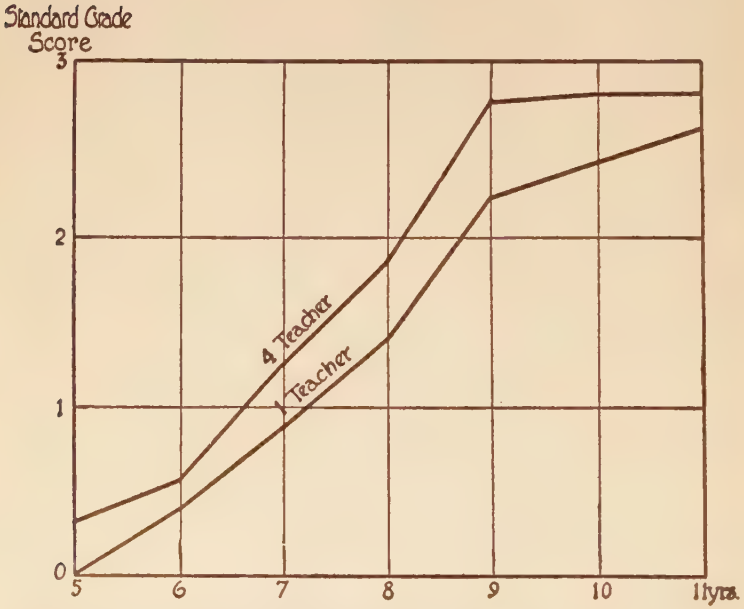


FIG. 15.—Increase in Reading by Ages in Grades 1-4 in Two Types of Schools. Adapted from Rural School Survey of New York.

B. Internal Factors

1. Breathing habits.
2. Vocalization and certain other features of inner speech.
3. Perceptual factors—absolute and relative perceptual spans.
4. Recognition and pronunciation of words.

5. Eye-movement habits.
   (a) Number and duration of fixations in oral and silent reading.
   (b) Number, duration, and type of refixations.
   (c) Return sweeps.
   (d) Interfixation movements.
6. Eye-voice span.
7. Meaning vocabulary.
8. Kinds of errors in oral reading and nature of each.
   (a) Mispronunciations.
   (b) Repetitions.
   (c) Insertions.
   (d) Omissions.
   (e) Substitutions.
9. Kinds of errors in silent reading.
   (a) Overpotency of words.
   (b) Neglect of key words and sentences.
   (c) Wrong relations.
10. Expression in oral reading.
   (a) Time.
   (b) Pitch.
   (c) Quality.
   (d) Force.
11. Appreciation.
   (a) Of the beautiful.
   (b) Of humor.
   (c) Of social values, including human nature.
   (d) Of intellectual or thought elements.
12. Rate of oral reading.
   (a) With various kinds of material.
   (b) According to the purposes for which it is done.
13. Rate of silent reading.
   (a) With various kinds of material.
   (b) According to purposes for which it is done.
      (1) Gist or general drift.
      (2) Details.
      (3) Answers to questions.
      (4) The most important ideas.
      (5) Careful organization of thought of selection.
14. Comprehension in oral reading.
15. Comprehension in silent reading, with reference to the items listed under 13 above.

From a study of this list it is readily seen that measurement of all the factors that affect reading achievement is a difficult task. The measurement of very many of these factors is not yet attempted except in a composite result. Diagnostic testing in this field is somewhat limited and inadequate. It is hoped that more effort may be given to this phase of the work.

Testing in the field of reading is inadequate not only because it fails to take into account and measure certain factors of reading efficiency but because the measures in use do not cover all the objectives of reading. The measures on the whole are measures of how efficiently people can read what they read. This phase of the reading program is basic and very important. However, one of our most desired objectives is "to develop strong motives for, and permanent interest in, reading." It is, of course, necessary that pupils be able to read well before this objective can be attained. If measurement is going to succeed, however, as an all around usable tool, it must give measures of all the objectives of a subject; reading is no exception.

The measures of reading are not so reliable as they should be. If it is expected that classification, promotion, and other administrative actions are to be taken on the results of reading tests, these should be more consistent in attaining the same type of result under various conditions. The tests of reading have usually been short, and this would tend to make the reliability low. Experimental comparisons of standard tests of various kinds have been carried on by Monroe, Ruch, Gates, and others. Results of this type of comparison are given in Table 22.

This table should be read as follows: Stanford, Form B, correlates .93 with a probable error of .010 with No. 1, which is Form A, Stanford Reading Test. Haggerty, Form A, correlates .88 with a probable error of .011 with Stanford, Form A, and .89 with No. 2, Stanford, Form B. In this manner the correlations between the various reading tests are given. The correlation between different forms of the same test is an indication of the reliability of the test.

TABLE 22.—CORRELATION BETWEEN READING TESTS AND BETWEEN FORMS OF THE SAME TEST[5]

(Grades IV to VIII. Number of cases, 154. 1926)

Reading Tests	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
1. Stanford A.....											
2. Stanford B.....	.93										
	.010										
3. Haggerty A.....	.88	.89									
	.011	.011									
4. Haggerty B.....	.88	.82	.83								
	.012	.017	.017								
5. Thorndike-McCall I.....	.80	.78	.71	.67							
	.19	.21	.026	.030							
6. Thorndike-McCall II.....	.71	.80	.68	.70	.75						
	.027	.019	.029	.027	.023						
7. Monroe 1.....	.78	.76	.79	.74	.48	.61					
	.021	.022	.019	.024	.041	.033					
8. Monroe 2.....	.76	.76	.75	.69	.53	.55	.76				
	.023	.023	.023	.028	.038	.037	.022				
9. Courtis 1.....	.61	.62	.58	.57	.51	.43	.60	.53			
10. Courtis 2.....	.64	.65	.61	.52	.59	.41	.52	.52	.74		
	.032	.031	.034	.038	.035	.044	.039	.039	.024		
11. Chapman	.68	.68	.69	.70	.59	.58	.58	.63	.53	.52	.89
	.029	.029	.028	.027	.035	.035	.036	.036	.038	.040	.011

[5] G. M. Ruch and W. F. Current, "Further Studies on the Reliability of Reading Tests," *Journal of Educational Psychology* (October, 1926), Vol. 17: pp. 446-481.

The studies of Monroe and Gates give somewhat different results. However, the results here given illustrate the point discussed. Gates correlated the scores on each test against a composite reading score obtained by use of several reading tests. He found, for example, that the Burgess Reading Test, although a short timed test, gives a correlation of .80 with this composite. This result indicates the value of the use of a good relatively short test for a rough measure, followed by an exhaustive retest and diagnosis of those cases that are revealed as especially in need of it.

The abilities involved in oral reading are somewhat different from those in silent reading. Not only are such abilities as clear enunciation, correct pronunciation, and control of breathing included in oral reading over and above those included in silent reading; but the mechanics of oral speech prevent development of speed beyond approximately 150 words a minute, while the silent reading rate may go as high as 400 or 500 words a minute.

The Gray Oral Reading Tests are an outgrowth of the Standardized Oral Reading Paragraphs by Gray.[6] These tests attempt to measure the mechanics of oral reading in terms of time and errors made and not in terms of comprehension or word meaning. If this test can be said to measure any objective, it is "to develop effective oral reading." The purpose and value of these tests are, however, primarily diagnostic. The manual of directions is fairly adequate. The reliability of the tests is not given. The norms are given by grades in terms of the rate and number of errors made. The scoring is objective. Teachers who use them should have some preliminary practice with them. These tests are made up of four sets, each of which has five forms of approximately equal difficulty. Set I is for grade 1, Set II for grades 2 and 3, Set III for grades 4 and 5, and Set IV for grades 6, 7, and 8.

[6] Published by the Public School Publishing Company.

The individual is tested alone with the teacher in a quiet place. He is given a copy of the test and told to "read this story to me. Begin reading when I say 'Begin.' Stop at the end of each paragraph until I say 'Next.' If you find some hard words, read them as well as you can without help and continue reading."

The following are instructions given the examiner:

In case a pupil hesitates several seconds on a difficult word, pronounce it for him and mark it as mispronounced.

While a pupil is reading, record on a separate test sheet the time required to read each paragraph and the errors made.

The time record is secured by noting the exact second at which the pupil begins reading a paragraph and the time when he completes it. The number of seconds required to read the paragraph should be recorded in the margin to the right of the paragraph.

The kinds of pupils' errors that are recorded are (1) total mispronunciation, (2) partial mispronunciation, (3) omission, (4) substitutions, (5) insertions, (6) repetitions.

Samples from Sets I, II, and IV are given. A part of Set IV is marked in the sample to show the method of marking.

GRAY ORAL READING CHECK TESTS[7]

Pupil's Name.....

N.B.—Make all records and notations below with pencil. Significant characteristics of the pupil's reading.

SET I—No. 2

A boy had a dog.
The dog ran away.

[7] Published by the Public School Publishing Company.

The boy ran after him.
He ran very fast.
He caught the dog.
He took him home.
The boy said,
"You are not a good dog.
You must stay at home."

SET II—No. 3

Our cow is red and white. Her name is Spot. There are red spots on her back. Her face and feet are white. She has four legs and a long tail. Her eyes are brown and very large. She has two long horns but we are not afraid of her.

SET IV—No. 4

many

The precipices rose like tower walls for ∧ thousands of feet. In the distance, the valley became narrow(er) until at last it dwindl(ed) to a gulch hardly thirty feet across, (but) of great depth. On the right rose a precipice-like wall of broken boulders that have had fallen from the cliffs.

In the part of Set IV shown, "precipices" was totally mispronounced, "narrower" was called "narrow" and thus partially mispronounced, "but" was omitted, "on the right" was repeated, the word "had" was called "have." In this way the pupil's errors are noted on the copy held by the teacher. An analysis of the number and types of error with a pupil or, if desired, with the class taken as a whole, will reveal specific difficulties and needs. The cause of an error is more significant, as a removal of the cause will correct the error. In the case shown, a poor enunciation of the type that omits final sounds or slurs them was probably the cause of the child's partial mispronunciation of "narrower" and "dwindled." This child must be taught to bring out final sounds of a word. To this end the teacher's own enunciation of word endings must be good so that imitation of a teacher's work will bring good results. It is well to develop enunciation exercises based on each pupil's needs. These should cause the opening of the mouth when speaking and the proper use of lips, teeth, and throat in word formation. The use of phonetics may be helpful in this respect. Special drill on endings such as *er, cd, ing, est,* and *ld* can be used with value.

A pupil's score or the score of a class may be compared to the norms given in Table 23. This comparison gives fuller meaning to the results of this test.

TABLE 23.—STANDARD SCORES

The standard midyear scores appear in the table which follows. The scores for rate and for accuracy are the same for each test of a given set. The columns headed R contain the rate records which indicate the total number of seconds required to read the selection of a test. The columns headed E contain the accuracy records which indicate the total number of errors made in a test.

Grades		*I*		*II*		*III*		*IV*		*V*		*VI*		*VII*		*VIII*	
	Set	R	E	R	E	R	E	R	E	R	E	R	E	R	E	R	E
Midyear	I	78	5	20	1												
	II			112	6	63	3	62	2								
	III							78	6	58	4	55	3				
	IV											69	6	64	5	60	4

Silent reading tests often attempt to measure separately two phases of reading—rate, the number of words read per unit of time; and comprehension, how much is understood. Some tests measure the degree of comprehension in a given time limit and allow this result to stand for reading ability or achievement. In measuring the rate of reading, whether silent or oral, the type of material used in the test has much to do with the score made. In measuring the degree of comprehension, it is necessary not only for the child to comprehend but to signify the degree of that comprehension by a supplementary performance that can be objectively scored. The performance that indicates his comprehension should be as objective and as controlled as possible.

The Gates Primary Reading Tests are a series of tests consisting of: Type 1, a test of word recognition; Type 2, a test

of word, phrase, and sentence reading; Type 3, a test of ability to read paragraphs of directions. These tests are scientifically developed and graded for difficulty. The content is validated in terms of scientific studies, better school practice, and the learning of children. The author of the series states that when "properly given, managed, and scored . . . these tests yield reasonably reliable results in the individual case and highly reliable scores for a class."

The manual of directions is excellent in most ways. It is especially helpful in discussion of causes of disability and methods of remedial instruction. The instructions for giving are on the test and not in the manual. The scoring is highly objective and can be rapidly done. The norms are given by age and grade for a representative group. Mental age reading norms are being worked out. Two forms of the test are available.

Type 1 is a test in word recognition. After the examiner has given the preliminary instructions, which include among other things the examiner's placing the drawings and words of the illustration on the board, the pupils are told "to look at the first picture. Next to it are some words. One of the words goes with the picture. You are to draw a ring around the one word that tells about the picture." This is then illustrated on the blackboard. The time is 15 minutes. The two illustrations on page 210 are taken from Type 1, Form 2.

Type 2 is a test of word, phrase, and sentence reading. In this type, as well as in Type 1, the examiner puts on the board a drawing of the illustrations used. After preliminary instructions, the pupils are told to "look at the words or story above the first row of pictures. Read what it says. Now find the picture below that tells about the words or story and put your finger on it." The examiner then illustrates the procedure. Certain suggestions concerning punctuation are given. Examples are given on page 210 from Form 1.

Type 1. Word Recognition [8]

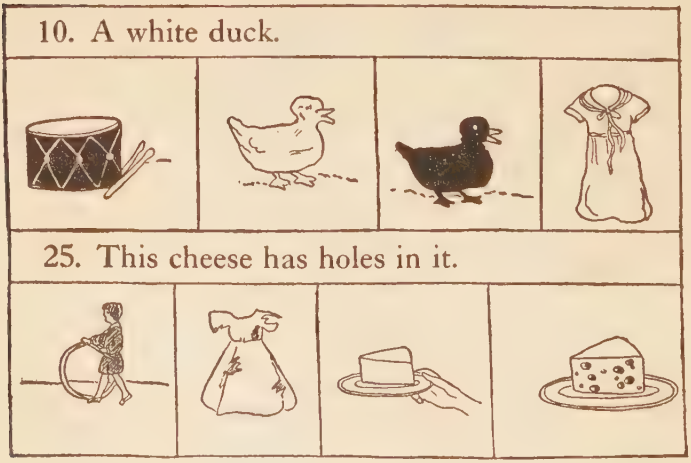
who two

own ten

comb camp

come lamb

Type 2. Word, Phrase and Sentence Reading



[8] From Gates Primary Reading Test, for Grades 1 and 2, published by the Bureau of Publications, Teachers College, Columbia University.

Type 3 is a test of the ability to read directions. The examiner uses the blackboard again for illustration purposes. The pupils are told: "We are going to see how well you can read. Do you see the pictures, and the stories below the pictures on the front page of your little book? Every one look at the first story and picture. Read the story to yourself." (Examiner points with finger to story on his own copy. Pause.) "What does it say to do?" (Have child read the story aloud to the class.) "That is right. Now read the story to yourself while I read it aloud." (Examiner reads slowly and clearly while children watch their own papers.) "Now watch what I am going to do." (Examiner makes mark called for on board with chalk.) "Read to yourself again what it says under the first picture. Then take your pencil and make the mark on your paper as I did on the board." Some samples from this test, Form 1, are given.

Type 3. Reading of Directions

10. Draw a line under the fat bear with the dish.

25. You must not cross the street when you see the word "Stop." You may cross the street when you see the word "Go." Make an X on the word that tells you it is time to cross the street.

The use that may be made of the results of these tests can best be illustrated by describing their use in an actual situation. Teacher X in a progressive city wished (1) to determine how well the pupils in her second grade class were progressing in

silent reading and (2) to determine the specific pupils that needed special attention and help. She gave Types 1, 2, and 3 of the Gates' Primary Reading Test. This skilled teacher applied the test carefully. The results are given in Table 24.

TABLE 24.—DISTRIBUTION OF SCORES MADE BY PUPILS ON GATES SILENT READING TEST

Type 1		*Type 2*		*Type 3*	
Score	*No. Pupils*	*Score*	*No. Pupils*	*Score*	*No. Pupils*
47-49	4	34-35	3	26	2
44-46	11	32-33	6	25	3
41-43	5	30-31	6	24	6
38-40	6	28-29	7	23	7
35-37	3	26-27	6	22	5
32-34	6	24-25	10	21	9
29-31	3	22-23	5	20	3
26-28	4	20-21	2	19	3
23-25	1	18-19	0	18	2
20-22	1	16-17	0	17	2
17-19	1	14-15	1	16	2
14-16	0	12-13	1	15	2
11-13	1	10-11	0	14	1
8-10	0	8- 9	0	13	0
5- 7	0	6- 7	0	12	0
2- 4	1	4- 5	0	11	0
0- 1	0	2- 3	0	10	0
No. pupils....	47	No. pupils....	47	No. pupils....	47
Class med. ...	39	Class med. ...	27	Class med. ...	21
Standard average for grade	33	Standard average for grade	22	Standard average for grade	22

By a study of Table 24 the standing of the class can be ascertained. This second grade class as a whole was achieving better than the norm; however, certain pupils fell far below the standard for the grade and needed special attention. The principal of this school was intensely interested in these results and gave special attention to the problem. Certain pupils were selected and grouped for study. Those who fell below 26 in Type 1 test, the test of word meaning, were studied to find their special difficulties. The study revealed that (a) these

pupils had a lack of confidence in attacking a new word or a word that they did not fully know; (b) they did not recognize likenesses in syllables such as "ing," "re," and "and," when found in other words; (c) they did not recognize, either in context or alone, words that the teacher had heard them use in talking. One boy of the group had an I.Q. of 83 and waited for the teacher's specific instruction before attempting anything.

Upon the basis of this meager analysis of difficulties remedial work was undertaken in this line. The material used for this work was found in another set of readers. The pupils were taught to attack a new word by recognizing familiar elements in it, by guessing its meaning in context, and by recognizing spoken words in written work. Informal tests were devised to determine whether the children really knew in print many of the words they used. To the surprise of the teacher it was found that they were totally unfamiliar with many words in print that they used in speech. These words were taught in a meaningful way, both alone and in context. A series of flash card games that made necessary the recognition of the words that were needed by these children was instituted. An ingenious teacher developed many variations of this flash card type of game. This type of drill, carried to what would seem an excess, was decidedly helpful. R.B., a boy who made a score of only three in the first test given in word meaning, showed the most development. Reading became to him a new joy that he had not known before. It might be added that he was saved from demotion and that he helped the teacher in thinking up new ways to play the game with flash cards. This pupil's trouble was primarily that he had not learned to associate spoken and printed words. He was also low in all phases of reading and took part in all the remedial work. His improvement was marked in all three types of reading tested.

Form 2 of these same tests was given to the whole class three months later. The average gain was slightly more than might be expected for the class as a whole. The average gain

of those cases that had had special remedial work in one or all phases of the work was marked. The lowest gain was made by pupil P.H., a boy with an I.Q. of 83, who gained approximately one-half year in reading ability as measured by this test. His greatest gain was in word meaning. The highest gain was made by pupil R.B., whose gain was as much as two and one-quarter years of normal progress.

The adverse criticisms of this test are primarily these: (1) the giving of the test is somewhat intricate due to the necessity of drawing and using the illustrations on the board; (2) the directions for the examiner are on the front page of the test; (3) some of the drawings are not as good as they might be, but perhaps they are, generally speaking, good enough. (4) Some pupils who are not through the second grade will make perfect scores on the test. This of course means that their reading ability is not measured. These are somewhat minor negative criticisms, and the weaknesses they indicate do not destroy the usefulness of the test.

Many reading tests are available for use in grades IV to the high school. It would not be possible here to discuss all of them in detail. The discussion here is limited to three tests that represent different types in form and function.

The Thorndike-McCall Reading Scales attempt to measure reading power in comprehension. The test is evidently not much of a measure of rate, since the time limit is 30 minutes. In this length of time many will complete the test. The selections of the test vary in difficulty from very easy to very difficult. A technical or special vocabulary is not found in even the hardest selections. There are about 20 exercises or selections in each test. The ability to comprehend each selection or exercise is measured by the ability to answer, in a short form, questions based on, or inferred from, the selections. A key of possible right and wrong answers is given in the manual. The directions and a preliminary exercise are on the front page of the test booklet. These directions and the exercise make

clear the method of doing the test. The manual of directions is valuable in the scoring, handling, and interpreting of the test scores. Nothing is given in the manual concerning diagnosis or remedial work, reliability, and validity. The reliability is not especially high. A correlation of reliability above .75 is not common for grades IV through VIII. The validity of this test as measured by its coefficient of correlation with a composite of other tests is $.73 \pm .12$. Norms are available by grade, age, T-score, and reading age. There are ten approximately equivalent forms of this test. This variety allows for repeated measurement with little practice effect.

The test can be best illustrated by giving two paragraphs from Form 3.[9]

Read this and then write the answers. Read it again if you need to.

In the spring men begin to work the soil. Those who have small gardens break up the ground with such tools as spades and forks. Those who live on farms turn the soil over with plows drawn by horses. In these ways the soil is loosened. After the planting is done, the weeds must be killed and the soil must be loosened. In places where little rain falls, other ways of watering the plants must be found.

13. What growing things does the gardener kill?.....
14. What animals draw the plows?.....
15. By whom is the plow used instead of a spade or fork?....
.....

Read this and then write the answers. Read it again if you need to.

On Monday Dick saw a red fox, a gray squirrel and a black snake in the woods. The next day he saw a brown rabbit and five brown mice in the field. He killed the fox and all the mice, but let the others live.

[9] Published by the Bureau of Publications, Teachers College, Columbia University.

16. Which animals did Dick kill?.....
17. Where did Dick see the rabbit?.....
18. How many animals did Dick kill?.....

This test has had wide usage. It is valuable for use in survey work to determine the proficiency attained by grades in comprehension of reading. It serves well as a good, general measure to determine results of experimental work or drill in comprehension of reading. It is useful in picking out for special study those pupils of a class or grade who are deficient in comprehension.

The scoring of the test is somewhat dependent upon subjective judgment, and is tedious. The defects may be overcome somewhat by practice in scoring it.

In the Courtis Silent Reading Tests (No. 2) an attempt is made to measure ability to read a story and to comprehend its content and the relationship of its elements. The test is made up of two parts. Part I measures rate of reading. This part is an interesting two-page story about "The Kitten Who Played May-Queen" (Form 1). The pupil is told to "read silently only as fast as you can get the meaning. . . ." At the end of each 30 seconds the examiner says "Mark" and the pupil draws a line around the last word read. Three minutes are allowed for the test. Part II is made up of questions on each paragraph of the selection. At the end of every five minutes the examiner says "Mark." The pupil draws a circle around the question on which he is working and goes on to the next set. The number of questions answered in the given time determines the rate of comprehension. Subtracting the number of wrong answers from the number of right answers, and dividing this difference by the number of right answers, gives an index of comprehension.

The test gives three types of scores: the rate score is rate of reading in terms of number of words read per minute; the rate of comprehension score is the number of questions correctly answered on the paragraphs; an index of comprehension

score is obtained by dividing the number of questions correctly answered into the difference between the number of right and the number of wrong answers. This index presupposes that the questions are of equal difficulty.

Instructions, Part I [10]

Do not open your paper until the examiner gives the signal to do so.

Directions to be read aloud by teacher and pupils together: This test is given to see how well you can read to yourself. When the signal to start is given, open the cover and begin to read the story. Read *silently,* and only as fast as you can get the meaning; for when you have finished, you will be asked to answer questions about what you have read. You will be marked for both how much you read and how well you understand it, but it is better to get the meaning of the story than to read too fast.

When the examiner says "Mark," draw a line around the last word read, and keep right on reading. If you should finish before the examiner says "Stop," close your paper and wait quietly until the others finish.

The examiner should allow exactly three minutes. He should say "Mark" at the end of each half-minute.

The Kitten Who Played May-Queen

5 When the spring sun was warm and the
13 spring flowers had begun to bloom, our children
21 had a May-party out on the lawn. Every little
30 boy or girl who lived nearby was invited. Betty
39 wrote the invitations and Bobby carried them to
47 the children.

49 When the day of the party came, Daddy planted
58 a May-pole and Mother tied it with gay-colored
66 ribbons. There were to be games and dances
74 on the grass and a delicious supper, with
82 a basket full of flowers for every child.

[10] From Form 1 of Silent Reading Test No. 2, of the Courtis Standard Research Tests, published by S. A. Courtis, 1807 East Grand Boulevard, Detroit, Michigan.

90 Under the big tree in front of the house
99 was a large box. Mother covered it with a
108 rug and put a little chair on it. This was
118 the throne on which the May-queen was to sit.

127 She was to wear a wreath of flowers, tied with
137 ribbons for a crown and over it a veil. Also, she
148 was to carry a beautiful May-basket and lead
156 the children in the dance around the May-pole.

164 "But who is to be queen?" asked the
172 children when all had come. "I am, for I
181 am the oldest child at the party," said a big
191 girl named Susan. "I vote for Susan," said her
200 chum Jane.

202 "No, Mary should be queen," spoke up one
210 of the boys, "for she is the prettiest girl."
219 But some of the other boys did not agree with
229 him. "Don't forget that Helen is the youngest,"
237 said Mother.

239 "Yes! Let's make Helen queen," cried Betty,
246 and all the rest of the children shouted, "Hurrah
255 for Queen Helen!" They carried her to the box
264 and placed her on the throne. Then Mother
272 put the crown and veil on her head.

Instructions, Part II

Do not open your paper until the examiner gives the signal to do so.

Directions to be read aloud by teachers and pupils together: Below will be found the first paragraph of the story, and under it five questions about the paragraph. After each question is a place to write the answer. Every question can be answered by "yes" or "no." Read the first question. If you can, write the answer at once. If you have forgotten the story, read as much of the paragraph just above the question as may be necessary to find the answer; then write the answer. In the same way, answer the other questions. When you have finished the five questions, wait for the other children to finish.

The examiner will then help you correct your paper and make sure you have understood directions.

THE KITTEN WHO PLAYED MAY-QUEEN

When the spring sun was warm and the spring flowers had begun to bloom, our children had a May-party out on the lawn. Every little boy or girl who lived nearby was invited. Betty wrote the invitations and Bobby carried them to the children.

A. Did the children have a May-party? .....
B. Was it Bobby who wrote the invitations? .....
C. Was the party held in the house? .....
D. Were only girls invited to the party? .....
E. Had the spring flowers begun to bloom? .....

The author of this test has made a distinct contribution in the method of interpreting the scores made by pupils of a certain grade on this as well as many other tests used under his direction. In Table 25 pupils *a, b, c,* and *d* are located by their score in the index of comprehension and by the number of questions answered.

The table shows that pupil *a* needs additional training in comprehension, but his rate of work is satisfactory. Pupil *b* is satisfactory both in rate of work and in comprehension. Pupil *c* is exceedingly poor (guesswork) in comprehension and is somewhat below the proper rate of work. Pupil *d* is satisfactory in comprehension but is in need of additional training in rate of work. Such a method of locating the pupil on a table in terms of his needs is a significant phase of accurate and meaningful measurement in education.

The Stone Narrative Reading Tests attempt to measure both rate of reading and ability to comprehend a narrative story. The test for grades III and IV is made up of two stories; for grade V, two stories; for grade VI, two stories; for junior high school, one story. After directions are read, a short preparatory test is given in which the details of the test are carefully explained. This explanation prepares the pupils for the story

TABLE 25.—FORM USED IN RECORDING THE SCORES OBTAINED FROM THE USE OF COURTIS' SILENT READING TEST NO. 2

Diagnosis		*Guesswork*			*Comprehension poor; additional training needed*					*Comprehension satisfactory*		
Questions Answered	*Total*	*Less than −5*	*−5 to +5*	*6–39*	*40–69*	*70–79*	*80–84*	*85–89*	*90–94*	*95–99*	*100*	
70												
65												
60												
55												
50								a			b	Rate of work satisfactory
45												
40												
35												
30				c								
25												
20										d		
15												
10												
5												
0												
TOTAL												

Diagnosis: Rate of reading too slow. Additional training needed.

Median Number of Last Question Answered.....
Median Index of Comprehension.....

which they are to read. The pupils read the two-page, interesting narrative story and at the end look up to find the number of minutes and seconds it has taken them to read it. They then turn to the next page and answer questions on the content just read. The questions are of the multiple response, single choice type.

An example from the preparatory test story of the seventh grade test is given.

Preparatory Test—Story [11]

A pot of gold was hidden at the end of the rainbow. One day a selfish man found it. He poured the gold pieces into a bag and ran to the woods to hide it. He did not know that there was a hole in the bag. As he ran, the gold pieces fell out and lay shining in the grass. There a little fairy found them. She did not want the selfish man to find them again. So she changed them into flowers for the children.

When the man missed the gold, he ran back to look for it. Not a piece of gold could he find. But bright golden buttercups—the first buttercups in the world—grew among the grass.

(The End)

Look up and get your rate score.

Copy your rate score on your answer sheet. Then turn this page to the questions on the other side.

The rate or number of words read per minute is recorded from a number on the time card exposed by the examiner every five seconds during the rate test. As each pupil finishes he looks up and copies from the time card the figure shown on it. The comprehension is determined by the number of questions answered correctly.

The measures of rate and comprehension given here may serve as a basis for picking out in the following manner those

[11] From Stone's Narrative Reading Tests, for Grade VI, published by the Public School Publishing Company.

who need help. Suppose that in the junior high school a test has been given to the eighth grade and that the results are available for each pupil by rate and comprehension.

TABLE 26.—DIAGNOSTIC RECORD CHART FOR STONE NARRATIVE READING TEST[12]

Make a Separate Sheet for Each Grade (Yearly or Half-Yearly)

City.....State.....Date.....Grade.....

School.....Teacher.....Section.....
(Upper or lower)

Rate	Comprehension																			Total
	1	2	3	4	5	6	7[a]	8[b]	9	10	11	12[a]	13[b]	14	15	16	17	18	19	20	
Over 500																					
480–499																					
460–479																					
440–459																					
420–439																1					
400–419									1												
380–399																					
360–379																					
340–359															1						
320–339					1					1	1										
300–319															1						
280–299[b]										2											
260–279[a]							1						1								
240–259											1	2									
220–239							1			3											
200–219													1								

[12] Published by the Public School Publishing Company.

Rate	Comprehension																			Total
	1	2	3	4	5	6	7[a]	8[b]	9	10	11	12[a]	13[b]	14	15	16	17	18	19	20	
180–199[b]																					
160–179[a]					1	1			1												
140–159																					
120–139 100–119																					
Below 100																					
TOTAL....																					

[a] For seventh grade make heavy rules as follows: vertically between 7 and 8 and between 12 and 13; horizontally between "260-279" and "280-299" and between "160-179" and "180-199." See directions next page.

[b] For eighth grade make heavy rules as follows: vertically between 8 and 9 and between 13 and 14; horizontally between "280-299" and "300-319" and between "180-199" and "200-219." See directions next page.

Medians: Rate..... Comprehension.....

Figure 16 has nine compartments in it. Each compartment is named. These refer to the diagnostic table (26) which has the pupils' achievement distributed on it.

High in rate Low in comprehension	High in rate Normal in comprehension	High in rate High in comprehension
Normal in rate Low in comprehension	Normal in rate and comprehension	Normal in rate High in comprehension
Low in rate Low in comprehension	Low in rate Normal in comprehension	Low in rate High in comprehension

FIG. 16.—Significance of the Divisions Shown in Table 26.

With this explanation of the table in mind, the teacher can pick out those pupils with outstanding needs. Certainly the

two in the compartment marked "Low in rate and comprehension" should receive first attention. The one pupil high in rate and low in comprehension should receive the next attention. In this manner remedial work can be given first to those who need it most.

Another particular type of test is exemplified by Gates Silent Reading Test, grades 3-8, and Sangren-Woody Reading Tests for grades 4-8. These tests attempt to measure certain particular phases of reading that represent special abilities or specific skills. These tests also measure, in a sense, the ability to read when the purpose is specific. It is true that people have need of reading in various ways and for various purposes. These tests attempt to reveal this differentiation in ability, skill, and purpose.

The Gates tests are four in number. Test *A* is designed to measure the skill of reading when the reader desires merely to get an accurate general impression or the "general drift" of what is read. A sample from the test is given.

Type A. Reading to Appreciate the General Significance [13]

This is to be a reading test. You are to read a number of paragraphs. Below each paragraph are five words. One of the words tells how some one described in the paragraph felt—whether sad or happy, etc. You should draw a line under that one—and only one—word to show that you understand just how the person described in the paragraph did feel.

> Once upon a time a young fairy went down to the river to swim. She jumped in with a splash. She put out her hands and tried hard to swim. Something seemed to be dragging her down. Oh, it was her wings! She had forgotten to take them off. Fairy wings become heavy when they are wet. She cried for help as loudly as she could.

[13] From the Gates Silent Reading Test, Form 1, published by the Bureau of Publications, Teachers College, Columbia University.

Draw a line under the word which tells how the fairy felt.

cross angry weary afraid joyful

Type *B* is designed to measure the ability to predict from an analysis of what is read the most likely event to happen next. The choice is made from possible events. An example from this test follows:

Type B. Reading to Predict the Outcome of Given Events

This is to be a reading test. You are to read a number of paragraphs. Below each paragraph are four sentences. Each sentence tells what is most likely to follow after the happenings that are described in the paragraph. You should draw a line under one—and only one—of these sentences to show that you can tell what will probably happen next.

The grocery man had a black cat. He loved his cat very much. One day a lady brought a big bulldog into the store. The grocer's cat raised his back and said "Meow! Psst!" to the bulldog. Of course, the dog did not like that, so he growled loudly. Before the grocery man or the lady knew what was happening, the bulldog had sprung upon the cat.

They let the fight go on.
The cat slept on.
The lady took her bird away.
The grocery man saved his cat.

Type *C* is designed to measure the ability to read exactly and precisely, and to pick out relevant details in what is read. This type is as different from Type *A* as the reading of the news of the daily newspaper is different from the reading of a recipe for the preparation of a cake. An example is given.

Type C. Reading to Understand Precise Directions

This is to be a reading test. You are to read a number of paragraphs. Each paragraph tells you to make some sort of

mark with your pencil to show that you have understood the paragraph. Do exactly what the paragraph tells you to do. Make the marks quickly—do not waste any time trying to make pretty drawings. The purpose of the test is to see how many of the paragraphs you can read in a short time. Don't waste any time. Don't look at any one else's paper. Remember, you must do exactly what the paragraph tells you to do. Don't make any marks other than those the paragraph tells you to make. Wait until you are told to "Begin," then turn the page and work as quickly and accurately as you can until you are told to "Stop!"

SUN	MON	TUES	WED	THURS	FRI	SAT

18. James and Mary go to a funny school. Instead of Saturday or Monday being their holiday, they have Wednesday and Sunday. They have to go to school on Saturday. Draw lines around the days of the week that James and Mary do not go to school.

Type *D* is designed to measure the ability to note details. This type is not so different from Type *C*, although it does not require such rigid and precise reading as does Type *C*. An example is given.

Type D. Reading to Note Details

This is a reading test. You are to read a number of paragraphs. Below each paragraph are three questions which you must answer by drawing a line under the one word or phrase which gives the best answer. Let us try a sample before we begin the real test. First read the paragraph. Then underline one—and only one—of the four answers to each of the questions to show that you understand what the paragraph said.

Next morning she awoke and found herself in a beautiful room. The walls were covered with silken

curtains. There were two mirrors made of pure silver. The bed was made of ivory. The coverings were made of silk and velvet. By her bed lay a dress and a pair of slippers. The dress was made of silk. The slippers were covered with diamonds.

Where did the girl find herself?
barn room garden store
What were the mirrors made of?
silver gold pearl silk
What were on the slippers?
rubies pearls opals diamonds

The Gates Silent Reading Tests are carefully constructed. The manual is valuable although not complete. The scoring is objective. Norms are available according to the various types of tests for various ages and grades.

The Sangren-Woody Reading Tests measure seven different phases of pupil's silent reading ability. They can be used in grades IV through VIII.

This test is composed of seven sub-tests. Test I is a completion test, testing word meaning. Test II is a story to be read silently. The rate of reading is tested each minute by drawing a ring around the word the pupil is reading when the tester says "Mark." Test III is a test on fact material. Test IV measures a pupil's ability to follow directions. Test V is a list of twelve exercises in which the pupil must get the total meaning. Test VI consists of ten short paragraphs followed by four questions, to test a pupil's ability to get the central thought. Test VII tests organization of material read. Two forms are available.

Tests of silent reading are primarily tests of objectives 3a "To develop well established fundamental reading habits" and 3b "To develop effective habits of intelligent interpretation." Although these objectives are measured in the various grades with different types of tests, they are the ones most nearly measured.

TESTS IN READING

The Chapman Unspeeded Reading Comprehension Test. By J. Crosby Chapman. Published: 1925. Purpose: to test pupils' reading comprehension. Range: grades 5-12. Time to give: 30 minutes. Time to score: 3 minutes. Standardized.
Publisher: J. B. Lippincott Co.
Price: $1 per package of 50.

Courtis Research Tests in Silent Reading, No. 2. By S. A. Courtis. Published: 1917. Purpose: to measure rate and comprehension in silent reading. Range: grades 2-6. Time to give: 8 minutes. Time to score: no information. Standardized.
Publisher: S. A. Courtis, 1807 East Grand Blvd., Detroit, Mich.
Price: $1.25 for material for 40 pupils.

Gates Primary Reading Tests. By Arthur I. Gates. Published: 1926. Purpose: to diagnose more elementary phases of reading. Range: grades 1 and 2. Time to give: 15 minutes. Time to score: short. Standardized.
The Primary Tests consist of three tests: Type 1, Recognition of Words; Type 2, Reading of Words, Phrases and Sentences, and Type 3, Reading of Paragraphs of Directions.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: $6 per 100 for 3 tests; $3 per 100, each test. Manual of Directions, 40 cents.

Gates Silent Reading Tests. By Arthur I. Gates. Published: 1926. Purpose: to yield a general measure of silent reading ability and to provide an intelligible diagnosis of special needs for follow-up work. Range: grades 3-8. Time to give: 30 minutes. Time to score: no information. Standardized.
The Grades 3 to 8 Tests consist of four tests: Type A, Reading to Appreciate the General Significance of a Paragraph; Type B, Reading to Predict the Outcome of Given Events; Type C, Reading to Understand Precise Directions, and Type D, Reading to Note Details.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: $8 per 100 for all four tests; $3 per 100, each test. Manual of Directions, 40 cents.

Gray's New Standardized Oral Reading Check Tests. By W. S. Gray. Published: 1923. Purpose: to determine the nature of the pupil's difficulties in oral reading and to measure the development of oral reading ability. Range: grades 1-9. Time

to give: approximately 2 minutes. Time to score: approximately ½ minute. Standardized.
Publisher: Public School Publishing Co.
Price: $1.50 for package of 100 tests—20 of each of the 5 tests of a check test set; 20 individual record sheets are furnished with each pad and 4 direction sheets; specimen set, 50 cents.

Haggerty Reading Examination. By M. E. Haggerty and others. Published: 1919, 1921. Purpose: to measure achievement and to aid in classifying pupils. Range: Sigma 1, grades 1-3; Sigma 3, grades 6-12. Time to give: no time limit. Time to score: short. Standardized.
Publisher: World Book Co.
Price: Sigma 1, $1.40 per package of 25 examination booklets, complete; Sigma 3, Form A or B, $1.50 per package of 25 examination booklets, complete; key for Sigma 1, 5 cents; key for Sigma 3, 10 cents; Manual of Directions, 30 cents extra; specimen set, complete, 50 cents.

Iowa Comprehension Test, Series B1, D1, and D2. By C. E. Seashore and others. Published: 1923, 1924. Purpose: to measure ability to comprehend what is read in the fields of history, literature, and science, at college level. Range: college freshmen. Time to give: 36 minutes. Time to score: 1 minute. Standardized.
Publisher: Extension Division, University of Iowa.
Price: $1.75 per 100 of either series, complete, plus transportation; specimen set, 10 cents postpaid.

Los Angeles Elementary Reading Tests: Form I. By Jessie E. Ingraham. Published: 1926. Purpose: to test a pupil's ability to get the idea which is conveyed by the material read. Range: grades 3-8. Time to give: 30 minutes. Time to score: short. Standardized.
Publisher: Research Service Co.
Price: $2 per 100 copies; specimen set, 10 cents.

Los Angeles Primary Reading Test: Form I. By Jessie E. Ingraham. Published: 1925. Purpose: to test a pupil's ability to listen, to read, and to think. Range: grades 1-3. Time to give: 10 minutes. Time to score: short. Standardized.
Publisher: Research Service Co.
Price: $1.50 per 100 copies; specimen set, 10 cents.

Minneapolis Primary Reading Test. By L. J. Brueckner. Published: 1923. Purpose: to test the pupils' understanding of words and meaning of sentences. Range: grade 1. Time to give: 30 minutes. Time to score: short. Standardized.
Publisher: Minneapolis Board of Education and Educational Test Bureau, Minneapolis, Minn.
Price: $2 per 100 copies; specimen set, 10 cents.

Monroe's Standardized Silent Reading Tests. By W. S. Monroe. Published: 1918; revised, 1920. Purpose: to measure rate and comprehension in reading. Range. Test I, grades 3-5; Test II, grades 6-8; Test III, grades 9-12. Time to give: 5 minutes. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: Tests I and II, 80 cents per 100; Test III, $1 per 100.

Pressey First Grade Reading: Attainment Scale Series. By L. W. Pressey and Viola Grant. Published: 1922. Purpose: to measure ability to recognize words and sentences. Range: grade 1. Time to give: about 10 minutes. Time to score: 30 seconds. Standardized
Publisher: Public School Publishing Co.
Price: 75 cents per 100.

Sangren-Woody Reading Tests. By Paul V. Sangren and Clifford Woody. Published: 1926. Purpose: to measure separately 7 different phases of pupils' silent reading ability. Range: grades 4-8. Time to give: 35 minutes. Time to score: 2 minutes. Standardized.
Publisher: World Book Company.
Price: $1.25 per package of 25.

The Stone Series of Narrative Reading Tests. By C. R. Stone and others. Published: 1922. Purpose: to measure rate and comprehension in narrative reading. Range: grades 3-9. Time to give: 30 to 40 minutes. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: 10 cents per copy or $7 per 100; specimen set, 40 cents complete; single set of time cards for 1 test, 75 cents; individual record sheets, 60 cents per 100; class record chart, 1 cent.

Thorndike-McCall Reading Scale. By E. L. Thorndike and W. A. McCall. Published: 1920. Purpose: to measure rate and comprehension. Range: grades 2-12. Time to give: 20 to 30 minutes. Time to score: 4 to 5 minutes. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: $2 per 100 each form; directions and record sheets are supplied with each order.

Van Wagenen Reading Scales. By M. J. Van Wagenen. Published: 1922. Purpose: to measure pupils' ability to read and comprehend English literature, history, or general science.

Range: grades 7-12. Time to give: about 50 minutes. Time to score: 1 minute. Standardized.
Publisher: Public School Publishing Co.
Price: $3 per 100, each series.

Williams Primary Reading Test. By Allan J. Williams. Published: 1926. Purpose: to measure ability to get the thought of the printed page. Range: grades 1-3. Time to give: 20 minutes. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: no information.

VOCABULARY

Armstrong-Danielson Sentence Vocabulary Test. By Caroline Armstrong and Cora Lee Danielson. Published: 1926. Purpose: to measure pupils' ability to recognize word meanings. Range: grades 3-9. Time to give: 15 minutes. Time to score: short. Standardized.
Publisher: Research Service Co.
Price: $1.50 per package of 100, Form I or II; specimen set, 10 cents.

Detroit Word Recognition Test. By Eliza Oglesby. Published: 1924. Purpose: to measure pupils' ability to recognize words. Range: grades 1 and 2. Time to give: 4 minutes. Time to score: 40 to 70 seconds. Standardized.
Publisher: World Book Co.
Price: 90 cents per package of 25 examination booklets, complete, Form A or Form B; specimen set, 15 cents postpaid.

Kennon Test of Literary Vocabulary. By Laura H. V. Kennon. Published: 1925. Purpose: to measure literary vocabulary. Range: teachers of English and undergraduate and graduate students of colleges and universities. Time to give: no information. Time to score: no information. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: $6 per 100, each form.

Pressey Reading Vocabulary Test for the Second, Third, and Fourth Grades. By L. W. Pressey. Published: 1921. Purpose: to measure the reading vocabulary of pupils. Range: grades 2-4. Time to give: 12 minutes. Time to score: 35 seconds. Standardized.
Publisher: Department of Psychology, Indiana University.
Price: 50 cents per 100, including directions.

Thorndike's Test of Word Knowledge. By E. L. Thorndike. Published: 1921. Purpose: to measure word knowledge. Range: grades 4-9. Time to give: 20 minutes. Time to score: short. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: $1.50 per 100.

BIBLIOGRAPHY

Averill, L. A. and Miller, A. D. "Size of Class and Reading Efficiency," *Elementary School Journal,* Vol. 25: pp. 682-91, May, 1925.

Ballenger, H. L. "A Comparative Study of the Vocabulary Content of Certain Standard Reading Tests," *Elementary School Journal,* Vol. 23: pp. 522-34, March, 1923.

Bonser, F. G. and others. "Vocabulary Tests as Measures of School Efficiency," *School and Society,* Vol. 2: pp. 713-18, November 13, 1915.

Breed, F. S. "Comparison of Two Methods of Measuring Comprehension in Reading," *School and Society,* Vol. 7: pp. 266-70, March 2, 1918.

Brinkley, S. G. "Relative Value of Different Types of Questions in Reading Tests," *School Science and Mathematics,* Vol. 25: pp. 703-09, October, 1925.

Brooks, Fowler D. "The Reliability of Silent Reading Tests," *School and Society,* Vol. 19: p. 652, May 31, 1924.

Brown, H. A. "The Measurement of the Efficiency of Instruction in Reading," *Elementary School Teacher,* Vol. 14: pp. 477-91, June, 1914.

Current, W. F. and Ruch, G. M. "Further Studies on the Reliability of Reading Tests," *Journal of Educational Psychology,* Vol. 17: pp. 476-81, October, 1926.

Dickinson, C. E. "Study of the Relation of Reading Ability to Scholastic Achievement," *School Review,* Vol. 33: pp. 616-26, October, 1925.

Elder, Harry E. "The Effect of the Summer Vacation on Silent Reading Ability in the Intermediate Grades," *Elementary School Journal,* Vol. 27: pp. 541-46, March, 1927.

Gates, Arthur I. *The Improvement of Reading,* The Macmillan Company, 1927.

Gates, Arthur I. "Series of Tests for the Measurement and Diagnosis of Reading Ability in Grades 3 to 8," *Teachers College Record,* Vol. 28: pp. 1-23, September, 1926.

Gates, Arthur I. "An Experimental and Statistical Study of Reading and Reading Tests," *Journal of Educational Psychology,* Vol. 12: pp. 303-14, 378-91, 445-64, September, October, November, 1921.

Goodenough, Florence L. "The Reading Tests of the Stanford Achievement Scales and Other Variables," *Journal of Educational Psychology,* Vol. 16: pp. 523-31, November, 1925.

Gray, W. S. "Case Studies of Reading Deficiencies in Junior High School," *Journal of Educational Research,* Vol. 10: pp. 132-40, September, 1924.

Gray, W. S. "Summary of Reading Investigations (July 1, 1924-June 30, 1925)," *Elementary School Journal,* Vol. 26: No. 6, pp. 449-59, February, 1926; No. 7, pp. 507-18, March, 1926.

Inglis, A. "A Vocabulary Test for High School and College Students," *English Leaflet,* Vol. 23: pp. 1-13, October, 1923.

Kelly, F. J. "The Kansas Silent Reading Tests," *Journal of Educational Psychology,* Vol. 7: pp. 63-80, February, 1916.

Kennedy, J. Alexander, and Carrigan, Emmet. "Comparative Study of the Results in Silent Reading in the Seventh and Eighth Grades in the Elementary and Intermediate Schools of Detroit," *Elementary School Journal,* Vol. 27: pp. 444-46, February, 1927.

Kyte, G. C. "Calibrating Reading Material by a Series of Tests for Comprehensibility," *Elementary School Journal,* Vol. 25: pp. 533-46, March, 1925.

McCall, W. A. and Crabbs, L. M. "Standard Test Lessons in Reading," *Teachers College Record,* Vol. 27: pp. 183-91, November, 1925.

Miles, D. H. "Significance of Reading in High School," *Contributions to Education,* Vol. 1: Chap. 25, World Book Company, 1924.

Neher, H. L. "Measuring the Vocabulary of High School Pupils," *School and Society,* Vol. 8: pp. 355-59, September 21, 1918.

Otis, Arthur S. "Consideration Concerning the Making of a Scale for the Measurement of Reading Ability," *Pedagogical Seminary,* Vol. 23: pp. 528-49, December, 1916.

Pressey, L. C. "Specific Elements Making for Proficiency in Silent Reading When General Intelligence is Constant," *School and Society,* Vol. 24: pp. 589-92, November, 1926.

Ritter, B. T. and Lofland, W. T. "The Relation Between Ability as Measured by Certain Standard Tests and the Ability Required in the Interpretation of Printed Matter Involving Reason," *The Elementary School Journal,* Vol. 24: pp. 529-46, March, 1924.

Thorndike, E. L. *The Teacher's Word Book,* 134 pp. Bureau of Publications, Teachers College, Columbia University, 1921.

Thorndike, E. L. "Word Knowledge in the Elementary School," *Teachers College Record,* Vol. 22: pp. 334-70, September, 1921.

"Tests to be Used in Measuring Appreciation of Literature," *School Review,* Vol. 33: pp. 491-92, September, 1925.

"Chronological Age as a Factor in Reading Achievement," *School and Society,* Vol. 20: pp. 826-28, December, 1924.

CHAPTER X

History and Civics

The social studies of the public school system are now in a stage of transition from special subject courses such as geography, history, and civics to an integrated or fused subject that is known as social science. This transition has met more general approval for the grades than it has for the junior and senior high school.

In the curriculum work done by the Department of Superintendence of the National Education Association, reported in the Third Yearbook of that department, social science and not the special subjects of history, civics, and geography receives the emphasis. More evidence of this unification is seen in the opinions of 148 schools in answer to the question,[1] "Do you favor the unification of the social studies into one course of study?" The answer was given for each grade. Table 27 shows the results of the answers in per cent favoring unification.

Table 27.—Per Cent of Replies by Grades Favoring Unification of Social Studies

Grades	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Per cent of replies favoring unification...	90.6	88.7	88.8	80.2	69.9	65.8	66.4	67.2	58.8	31.3	31.9	37.7

This report, however, represents the desires of educators and not the general practice, for as yet the conventional subject divisions of history, civics, and geography are largely retained.

[1] This study and report made by A. C. Senour, General Supervisor, East Chicago, Indiana, schools and reported in *Fourth Yearbook*, Department of Superintendence, p. 327.

The junior high school, however, will probably come rapidly to the point of view of a unified or integrated course. This is doubtless due in large part to a desire to build the social science work around the social problems that confront humanity. This requires a study of the social, economic, historical, political, and geographic influences that bear on the solution of the problems studied. Such a method of organization involves the inclusion of much material not generally given in separate courses of history, civics, and geography. In many of these courses there are elements of likeness in content due to similarity of aims or objectives. If the subjects are taught as special separate subjects, it is very difficult to devise a time schedule that would make the elements of these special subjects helpful to each other. In the second place, the special subject teachers of history, geography, civics, and other studies do not find it easy to correlate and coördinate their work without overlapping. In many cases these teachers are trained for high school work and do not find it easy to teach beyond the logical arrangement of their special subject matter. An integrated course in which the lines of demarcation between subjects are not drawn is more likely to recognize, in the organization and presentation of the subject matter, the applications of the principles of the new psychology of the junior high school pupil. Such an integrated course is more likely to coordinate its material with the life the child leads in school and the demands of society outside the school.

Testing in the integrated social science lags far behind the development of this newer idea of curriculum making in this field. There is a very great need of standard tests in this newly organized subject. Such tests could be developed on the basis of the objectives set forth to be attained in such a course and in accordance with the best courses of study now in use in this field.

The Kepner Background Tests in Social Science[2] should

[2] Published by Ginn and Company.

receive recognition as a good start in the direction of social science testing in the integrated field. These tests are primarily inventory tests. They "aim to measure the extent of the factual background in the social sciences which a pupil or group of pupils bring to any one of the social science subjects in the secondary school." The tests are of value primarily as diagnostic or inventory measures of content that it is assumed a beginning ninth grade pupil has learned in the fields of history, civics, economics, geography, and sociology. The emphasis on the content chosen for the test is in the order named in the previous sentence. The author made an analysis of the "twenty-one representative textbooks" to determine the assumed background for the subjects represented by these textbooks. After a critical analysis by teachers and an experimental tryout of the test with 9th, 11th, and 12th grade pupils, the test was published. Exercise I of the test is a test of association of men-events and events-men, examples of which are:

Exercise I

In each of the sentences below you are given four choices for the last word. Only one is the correct word. In each sentence draw a line under the correct word.

3. A foreigner helping the Americans during the Revolutionary War was
   Webster Lafayette Newton Genêt

11. Down to 1924 the last Democratic president was
   Cleveland Harding Taft Wilson

Exercise II is a test of literary background. Examples follow.

Exercise II

In each of the sentences below you are given four choices for the last word. Only one is the correct word. In each

sentence draw a line under the correct word as you did in Exercise I.

6. "The Leather Stocking Tales" tell of life among the
Vikings Crusaders Russian Cossacks Indians

7. "The Tales of King Arthur" tell about the
Feudal Age Crusades Renaissance Reformation

Exercise III is a completion test of geographic concepts.

Exercise III

In the sentences below one and sometimes two words are omitted at the end of the sentence. In each case write the proper word or words in the blank space or spaces so as to complete the sentences accurately.

4. The thirteen colonies in America were located along the .....
13. Distance on the earth's surface measured east and west is called .....

Exercise IV is a best answers test of historical words and terms commonly used.

Exercise IV

Examine carefully the words and phrases that follow and indicate the correct phrase with a mark X.

5. PROPAGANDA is a method of spreading
....germs of deadly diseases
....wireless messages
....beliefs, opinions, or principles
11. The RENAISSANCE was a medieval movement marking a
....renewal of interest in art, literature, and learning
....struggle between nobility and common people
....struggle between England and France

Exercise V is a single word answer test on civic, social, and economic terms.

Exercise V

In this exercise read the descriptions carefully and in the list of words given below, find the word which is described. When you have found the word, write it in the blank space opposite the description.

3. Making war on your country or helping an enemy of your country. .....
11. That part of the government determining the meaning of laws. .....

4. Judiciary	14. Cabinet
5. Political Party	15. Treason

Exercise VI is a test of chronology of men or events.

Exercise VI

In the following exercises indicate the time order in which each took place (held office, lived, etc.) by placing the figure 1 before the event which took place first, the figure 2 before the event which took place second, and so on through the five events.

2. Presidents
....Roosevelt
....Jefferson
....Lincoln
....Wilson
....Washington

4. Events
....Napoleonic Wars
....Slavery Agitation in America
....Feudalism
....Fall of Rome
....Discovery of America

Exercise VII is a matching test of dates and events.

Exercise VII

In the following exercise a list of dates is given at one side of each exercise, and at the other side there are blank spaces. A date is to be taken from one side and put in the proper blank space at the other side for each event.

2. EVENTS

1863	Declaration of Independence	.....
1492	Emancipation Proclamation	.....
1765	Discovery of America	.....
1620	Settlement of Plymouth	.....
1776	Stamp Act	.....

5. REPUBLICS (When established)

1919	China	.....
1848	United States	.....
1912	France	.....
1792	Germany	.....
1789	Switzerland	.....

The test is not a time test. The manual of directions is helpful. The scoring is objective and easily done. The reliability is only fairly high, .79 for high school juniors and seniors. There are two adverse criticisms of this test.

(1) Its content is so much historical that its measurement is primarily of historical knowledge. This is shown by its correlation with the history test in Table 28. (2) The value of the test as an inventory or diagnostic test is lessened by the inclusion of so few items and consequently such a small part of the assumed background, not only of facts but of concepts as well. Notwithstanding these criticisms, the work of the test has been carefully done and is a step in the right direction.

As desirable as it is that integrated social science be taught and tested, nevertheless most of the present social science tests are in the fields of special subjects. It is therefore necessary to discuss the subjects and the tests in these subjects separately and without reference to their coördination, fusion, or integration with each other.

HISTORY

History has an important place in the curriculum of our public school system as measured both by time given to it in

the daily program and by the interest shown in it by educators. Interest in history is often the result of a belief that a study of the past will give guidance for action in the future. Although changes have taken place in the various statements of aims of the teaching of history, this one belief has always prevailed. A recent statement of the objectives of the teaching of history gives the following:[3]

1. To develop the ability and habit of acting as an intelligent member of a human group and especially of our national group. This involves:
   a. An active loyalty to the ideals, standards, and faiths of the United States.
   b. A participation in the thoughts and actions of our country.
   c. An intelligent appreciation of the relation of the individual to the country.

2. To develop an intelligent appreciation of the social, economic, political world in which we live by:
   a. Learning to know, appreciate, and make use of the contributions of the people of the past of our nation and the nations and races of other lands.
   b. Learning to know the laws and principles underlying human progress and to conserve these human values which have been discovered.
   c. Learning to think of social problems in terms of history.

3. To develop a sense of ethical values.

Lack of agreement concerning the objectives of instruction in history is marked especially by the difference in emphasis given to certain phases of it by different teachings. Not enough analysis of the subject into the desired knowledge, skills, habits, and attitudes has been made to allow for careful curriculum

[3] Stratemeyer and Bruner, *Rating Elementary School Courses of Study*, Teachers College, Columbia University, 1926.
Arthur S. Gist, *Elementary School Supervision*, Charles Scribners' Sons, 1926.
Paul Klapper, *The Teaching of History*, D. Appleton and Co., 1926.

construction and testing. Minimum essentials are not adequately agreed upon and stated so as to make progress from grade to grade a scientific procedure.

A basic knowledge in history is the ability to retain a few dates exactly, more dates in approximation, and time sequence in many things. It is true that this information is useless except when its meaning is fully associated and interpreted and made to function in the life of a child, in his reading, thinking, and expression. Nevertheless these bits of information are basic, and a measure of them is justified. The real deficiency in tests is not that these facts have been included but that other abilities more important to the attainment of the objectives of history instruction have not also been included.

History tests that attempt to measure the same thing seem to give very different results. The table given illustrates this fact by giving the correlation coefficient between various tests of history.[4]

TABLE 28.—INTERCORRELATION OF TEN TEST FORMS OF AMERICAN HISTORY TESTS

Test Form	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*Average*
1. Gregory A	..	.79	.59	.69	.72	.68	.76	.71	.49	.51	.66
2. Gregory B	.79	..	.52	.61	.72	.70	.72	.69	.49	.45	.63
3. Barr A	.59	.52	..	.71	.61	.52	.60	.53	.35	.44	.54
4. Barr B	.69	.61	.71	..	.69	.57	.65	.56	.53	.52	.61
5. Pressey and Richards	.72	.72	.61	.69	..	.67	.82	.71	.56	.55	.67
6. Van Wagenen S-3	.68	.70	.52	.57	.67	..	.72	.66	.45	.44	.60
7. Kepner A	.76	.72	.60	.65	.82	.72	..	.79	.54	.47	.67
8. Kepner B	.71	.69	.53	.56	.71	.66	.79	..	.38	.40	.60
9. Van Wagenen History Reasoning A	.49	.49	.35	.53	.56	.45	.54	.38	..	.57	.48
10. Van Wagenen History Reasoning B	.51	.45	.44	.52	.55	.44	.47	.40	.57	..	.48

This table should be read as follows: Gregory History Test, Form A, correlates .79 with Gregory History Test, Form B; with Barr History Test, Form A, .59; and in the same way

[4] G. M. Ruch and others, *Objective Examination Methods in the Social Studies.* Published by Scott, Foresman and Company, 1926.

with all the other seven test forms given. In the column under "average" is given the average correlation of this test with the other nine forms. In the case of Gregory History Test, Form A, this is .66.

The correlations between the various history tests given in the preceding table are somewhat low. Such low correlations between tests, some of which purport to measure the same abilities, indicate a lack of consistency in the field of history testing. The lack of consistency is probably due to a lack of common content and emphasis in these tests.

American History

The Van Wagenen American History Scales [5] originally attempted to measure separately, (a) amount of knowledge, the test for which was called Information Scale; (b) the ability to draw inference from historical data, the test for which was called Thought Scale; (c) the ability to judge character and motives from facts about persons, especially in their public life, the test for which was called Character Judgment Scale. In the more recent revision of these scales the last one has been omitted.

These scales are now made up of 11 scales, including Thought Scale R2 in General History; Information Scales S1 and S2 in General History, and their alternate forms, R1 and R2; Information Scales C1 and C2, covering the period from the discovery of America to the Revolutionary War; Information Scales F1 and F2, covering the period from the Revolutionary War to the Civil War; and Information Scales K1 and K2, covering the period from the Civil War to the present time. These scales supplant the Van Wagenen History Scales—Information and Character-Judging.

As examples of these tests the Information Scale K2, a test

[5] Published by the Bureau of Publications, Teachers College, Columbia University.

on the period from the Civil War to the present, and the Thought Scale R2 are here described.

Information Scale K2 is composed of 30 questions, some questions containing more than one item. The questions are arranged in three groups of increasing difficulty, each group being given a difficulty rating. The questions are in type much like those of the classroom. The child is told to "begin with question 1 and answer as many questions as you can. Answer them in order. If you cannot answer all the parts of any question answer as many parts as you can. Start."

The following examples are taken from the test.

Group I (Average Value 68.5)

1. (*64*) After each of the movements write which one of these men—Thomas A. Edison, William J. Bryan, the Wright Brothers, Woodrow Wilson, Jefferson Davis—took a leading part.

Southern Confederacy:
Invention of the aeroplane:
Free silver movement:

9. (*72*) Of these five events, put a check mark in front of the two which happened about the same time.

....Venezuelan dispute
....Establishment of the Civil Service Commission
....Spanish-American War
....World War
....Purchase of Porto Rico

Group II (Average value 78.5)

11. (*74*) Put a check in front of each of these things which happened during President Roosevelt's administration.

....Spanish-American War
....Haymarket riots
....Panama Canal construction
....Underwood tariff
....Venezuelan controversy

22. (*84*) What are the two main purposes of any tariff act?
1.
2.

Parts of Thought Scale R2 for grades seven and eight are also given.

DIRECTIONS

In each of the problems below you will find one or more short paragraphs giving you certain facts. Below the paragraphs are questions or statements telling you to do something. This is not a test to see if you remember something which you have learned. Read the paragraphs carefully and from the facts given in them think out the answers to the problems for yourself. Begin with problem 1 and do as many as you can. Do them in order. If you cannot do all the parts of any problem do as many of the parts as you can. Start.

Group I (Average value 70)

4. (*70*) In 1793, Eli Whitney invented the cotton gin, a machine for separating the cotton seed from the fiber. By the use of this machine one slave could clean fifty times as much cotton in a day as with the old machines or by hand.
What effect would this invention have upon the amount of cotton raised?

Group III (Average value 90)

27. (*89*) We know that in 1790 there were about 3,929,200 people living in the United States. We do not know at all accurately how many people there were in the colonies at any previous year after 1650.
Why do we know how many people there were in the United States in 1790, but do not know how many there were in any year before that?

All of the keys for scoring answers to the Van Wagenen Scales allow one-half credit for answers half correct. The key is specific in the directions for correcting the papers. The

scoring of the test is such as to give an accurate measure of a pupil's achievement. The method of finding the final score is very intricate. The difficulty standard is 75, which is based on the average achievement of twelve-year-old children. All scores as finally arrived at are comparable to this difficulty standard of 75. This is of value especially in that differences are in equal terms so that a pupil making a score of 65 is as much above one making 60 as one making 95 is above one making 90. This is not true of most tests due to the fact that many of the scores (if used as in raw form) are clustered around the center of the distribution. The method that Van Wagenen uses makes a yardstick out of the scale. Norms are available by chronological age, grade, T-score, and mental age.

These tests represent careful and painstaking work not only in the technique of development but also in the choice of material, for the tests represent acceptable items and cover a good range of essential elements of American History. The user of the test may, however, desire a more objective type of test which could be made by changing the type of question to a controlled multiple response, rather than allowing the freedom in type of answer that the present form of test permits. The scoring of the test scales is intricate, and this objection will limit practical use of the scale in the schoolroom.

A difficulty encountered in the measurement of "ability to draw inferences from historical data" is suggested by the Thought Scale R2. Question 27 (see page 245) might possibly be answered correctly because of the child's remembering that in 1790 the first census was taken. In such case the thought question becomes a question that merely tests memory and not "ability to draw inferences from historical data." This weakness, however, does not destroy the value of this test as a pioneer in the type of testing represented by it.

Some of the characteristics of other tests in American history are worthy of comment and illustration. The Gregory Tests of American History, Test I and Test II, which cover

the important phases of the history of the United States, can be used with some degree of satisfaction in the seventh and eighth grades before the entire completion of the history of the United States. The development of the two tests covering separate units of history makes possible classification and adjustment to needs before the work is completed. It gives a chance for analysis of achievement because the parts of the test are organized upon the basis of units of the courses of study; for example, Part II covers the period of expansion and discovery. The tests are highly objective and easily scored.

The Barr Diagnostic Tests [6] in their title are suggestive of what they attempt to measure. The first part, Test 1, is an attempt to determine how well the pupil can read and understand historical material with an historical vocabulary. An example is given.

Test I

1. Read through the following paragraph rapidly. Answer as many as you can of the questions which follow. Read the paragraph again if you need to.

> In 1777 the British planned a three sided attack upon the colonies. Colonel St. Leger was to march east from Lake Ontario and Oswego along the Mohawk River in the State of New York. General Burgoyne was to come south from Canada by the way of Lake Champlain and General Howe, or at least a portion of his army, was to move north from New York City along the Hudson. The three armies were to meet in the vicinity of Albany. St. Leger attacked Fort Schuyler on the west but failed and returned to Canada. Howe did not receive his final instructions from England until he had moved in an opposite direction, and therefore too late for him to give assistance. Burgoyne proceeded to the southward as planned but was completely defeated and was forced to surrender at Saratoga October 17, 1777. The victory was not only the turning point of the war but one of the "decisive battles of the world."

[6] Published by the Public School Publishing Company.

(a) Put a cross (X) before each of the following statements that is discussed in the above paragraph.
....(1) The French and Indian wars.
....(2) The British campaign of 1777.
....(3) The battle of Saratoga.

(b) Draw a line under one of the following words which shows how many military expeditions were planned by the English.
Two Three Four

(c) Draw a line under the name of one of the following men who failed to take part in the triangular campaign.
St. Leger General Howe
General Burgoyne

(d) Draw a line under one of the following words which tells where the British armies were to meet.
New York City Lake Champlain
Albany

(e) Draw a line under the most important event in the above paragraph.

Test II is a test of chronological judgment such as is usually included in a history test.

Test III is an attempt to test the ability of a pupil to judge the value of types of historical evidence or historical material in history. An example is given.

Test III

1. Read through the reading passages hurriedly. Concentrate upon the questions and directions.

If you were writing an account of an early Indian raid and came to a point in the story where there were the three following conflicting accounts, which of the accounts (from all you can learn from the accounts themselves) would you accept as probably the correct version of the story?

Put a figure one (1) before the account which you think

is most likely to be the correct one; a figure two (2) before the one that is next in value; and a figure three (3) before the one that is least likely to be the correct version.

....(a) Mr. Johnson, an old settler living in the neighborhood, said that in his opinion the Indians were all killed before crossing the river.

....(b) Copies of old letters from Mr. Smith to his wife and Mr. Hale to his son, say that only one Indian was killed and that the rest escaped across the river to the southward.

....(c) Mr. Meeks, another old settler, who was an eye-witness, said that the Indians were scattered but escaped across the river without loss.

Test IV attempts to measure the pupil's knowledge of the importance of facts or events.

Test IV

8. Put a cross (X) before the fact in the following list which has been of the greatest importance in the agricultural development of the United States.

....(a) The invention of the McCormick reaper.
....(b) The Federal Farm Loan Act.
....(c) The application of machinery to farming.
....(d) The Lincoln-Douglas debates.
....(e) The completion of the Roosevelt Dam.

Test V is an attempt to measure the pupil's ability to evaluate and associate the facts given in a paragraph. An example is given.

Test V

3. In the left hand (first) column below is a list of causes. In the right hand (second) column is a list of results. Take your pencil and draw a line connecting each cause with the proper result.

(a) The growth of industrial combinations	(a) The Mexican War
(b) The War of 1812	(b) The extension of slavery
(c) The spread of cotton culture	(c) The early westward movement

(d) The Homestead Act (1862)	(d) The Interstate Commerce Act (1887)
(e) The annexation of Texas	(e) The development of farming in the far West

If the parts of this test were lengthened to the point where the results would be a reliable measure of each of these abilities and if the test were well standardized and made more objective in scoring and perhaps extended to include a test of the ability to apply historical facts to present-day problems, it would take the front rank in history tests for secondary schools. The test is not diagnostic in the usual sense of that word.

The test should be very suggestive to teachers of history in that it measures those abilities which are often entirely omitted from measurement by other history tests and especially by the traditional written examinations.

The Pressey-Richards Test for Understanding of American History has one feature that may suggest some diagnostic significance in this type of test. The second part, Test *2*, of this test is an attempt to measure knowledge of historical vocabulary. The words included here are those found to be important in the study of technical vocabularies by Pressey.[7]

The technical vocabulary of a subject is of importance as a measure of ability to read the subject and to think in terms of it. A diagnostic study of the pupil's knowledge of a rather large sampling of basic vocabulary of history might reveal a specific difficulty that would be easily remedied. An example of Test *2* is given.[8]

Test 2: Historical Vocabulary

Directions: On the page below there are some questions. Each question is followed by four answers; only one of these

[7] *The Technical Vocabularies of the Public School Subjects,* Public School Publishing Company.

[8] Pressey and Richards, "A Test for Understanding of American History," Public School Publishing Company.

answers is right. You are to find the right answer and draw a line under it. The first question has been correctly answered.

1. What is a stockade? A part of a gun A supply of ammunition A public punishment A barrier for defense.
2. What is an envoy? A transport Spy Diplomatic agent Bill.
3. What is a dynasty? A ruling family An earthquake An electrical appliance A league of nations.
4. What is a confederacy? A disunion A colony A commonwealth A league of states.
5. What is an autocracy? Representative government Mob law Self-government An absolute form of government.
6. What is meant by secession? Withdrawal from a union Accession to a throne A riot A civil war.
7. What is a referendum? Power to initiate legislation A veto Reference of legislation by the people Return of a bill without signature.
8. What is a maneuver? A political campaign A mutiny A declaration of war A movement of troops.
9. What is an evacuation? A surrender An invasion Lack of national funds A complete withdrawal of troops.
10. What is a galleon? A sailing vessel A Spanish nobleman A piece of money A battering ram.

The whole test is objective and easily scored. The reliability of the test is higher than that of other history tests (see Table 28). Its correlation with other history tests is relatively high.

The Diagnostic Test in Modern European History, by C. G. Vannest, was published in 1921. This test is designed to diagnose high school pupils' difficulties in history. It takes 40 minutes to give and approximately 5 minutes to score. It is standardized tentatively. The test consists of five different tests: (A) Time sense, (B) Place sense, (C) Evaluation of facts, (D) Thought, (E) Information. Each test is composed of several exercises covering the general field of modern Euro-

pean history. Only one form is available. The content of the test is based primarily on Robinson and Beard's *Outlines of European History,* Part II.

Little help can be obtained from history tests for diagnostic work. However, the results of the Gregory test, for example, can be used for classification, for teaching purposes, and for pointing out some specific needs of the group on particular phases of the work. In connection with whatever diagnostic and remedial work is done in this field one must keep in mind at least three phases of history teaching. In the first place, history has a factual side that is important and basic. Inadequacy of knowledge of minimum facts is as serious in history as inability to add the higher decade combinations is in arithmetic.

It must be remembered, however, that these minimum facts are not the end but only a means to the end of building up ability to solve problems in the social world. It is in just the same manner that the ability to bridge the tens in higher decade addition is only a means to the end of solving problems in arithmetic. The acquisition of these facts of history so that they become as automatic as 7×3 is dependent upon review, repetition, and drill, for which there is little or no provision in the history work of the public schools. The solving of problems, the solution of situations, becomes an important phase of the teaching of history. Problem solving in history, just as in arithmetic, has many phases. An understanding of the problem at hand means reading the problem to comprehend it, picking out from the facts given those that are pertinent to the problem, deciding upon a method of solution, testing the results, etc. If the problem is not solved, the difficulties may be ascertained by analysis of the pupil's thinking. In this manner real diagnosis in history is possible, although not attained in the slightest degree as yet.

Another of the phases of history teaching so much spoken of in the objectives set out for it and given so little attention in the teaching of history is the application of the principles and

methods of solving problems of the past to the problems of today and to the problems of the next few years. If history is "to develop the ability and habit of acting as an intelligent member of a human group and, especially, of our national group," then the methods and principles of solving problems in history must be transferred to life situations of the present and probable future. Diagnostic testing of a kind that will measure progress in this line and point out specific needs in teaching is badly needed.

Much of the value of history teaching comes through the development of the emotions. Proper development of emotions means building up of the proper attitudes. It is on the basis of attitude or emotional reaction that much of the action of individuals depends. Testing ability to solve problems and attitudes developed as a result of history teaching is very much needed.

An effort has been made by Odell[9] to develop scales of answers to certain thought questions in various social science subjects. These questions in the American History Scale are of nine types—Type I, Analysis; Type II, Cause and Effect; Type III, Comparison; Type IV, Criticism; Type V, Discussion; Type VI, Explanation; Type VII, Relationship; Type VIII, Reorganization; Type IX, Summary. The scales are like those of composition in form. The one discussed here is in American History. The answers to the questions used in the scale are taken from the answers made by pupils. The answers have been rated from 0 to 10, the poorest answer being rated 0 and the highest 10. Following the answer given in the scale, a criticism is made of each answer showing its good points and its deficiencies. The scale should be helpful to teachers in standardizing the grading of answers to thought questions. Examples are given.

[9] "Scales for Rating Pupils' Answers to Nine Types of Thought Questions in American History," Bureau of Educational Research, University of Illinois, 1927.

Type III.—Comparison

Compare the policies of the Republican and Democratic parties since the Civil War.

VALUE 1

Pupil Answer. Before the war the policies of the Republican party and the Democratic party was slavery. But after the Civil War every thing is in destruction in the South. The main policies of the parties was the reconstruction of the South. Some of the main problems was; shall the negro have the rights of an American citizen? Shall all the negroes vote or only a part of them? And such questions as these confronted them.

Criticism. Apparently the writer of this had some understanding of what the question meant in regard to the word policies. Some effort is made to discuss reconstruction as an issue, but nothing else is treated. The first sentence refers to policies before the Civil War not called for by the question and the questions at the end are hardly relevant.

VALUE 5

Pupil Answer. Since the Civil War the Republican party has been in power most of the time. Soon after the close of the war the Republican party stood for reform. In Grant's first administration several reforms were made and many gangs unearthed as the Tweed Ring, Whiskey Ring, and Credit Noluliere's. Several other reforms were made in the civil revenue. Their political platform was reform. The Democratic party only criticized the Republican party and tried to make a platform out of that.

The Tariff question caused a background for both platforms. The Democrats were in favor of a low tariff and the Republicans in favor of a high Tariff. The silver question was favored by Democrats and disapproved of by the Republicans. The World War was also another platform.

Criticism. Policies of reconstruction are discussed and the World War spoken of though really not treated. The general matter of reforms is also introduced, but trusts are omitted and the tariff and money issues treated rather indefinitely. The attitudes of the parties of reconstruction are not brought out.

There are in all four points of comparison mentioned in this answer.

VALUE 10

Pupil Answer. For some time the Democrats favored the free coinage of silver at a ratio of sixteen to one, while the Republicans were in favor of gold as the single monetary standard. The Republican party favors a high protective tariff, and the Democratic party favors a low tariff with many additional items on the free list. In the more recent platforms we find the Republicans praising the immigration laws and favoring improved naturalization laws, while the Democrats favor the exclusion of Asiatic immigrants. The Republicans favor universal mobilization in war time, while the Democrats favor the drafting of resources and a referendum on war unless attacked. Republicans favor the continuation of our present policy in the Philippines, while Democrats favor immediate independence. The Republicans favor the establishment of a Permanent Court of International Justice, while the Democrats stand for no substitute for the League of Nations.

Criticism. Among the additional points mentioned in this are the immigration question, the attitude of the two parties toward the Philippines, the mobilization of resources as favored by the Republicans, Democratic endorsement of drafting resources, and a referendum on war unless attacked.

Civics

The two following quotations concerning civics are those of thinkers in that field, and summarize briefly the material and objectives of this subject.

Knowlton [10] defines civics thus:

> The core of civics should be government or what is commonly called political science. It begins with government and ends with government, no matter what excursions may be taken from it into other realms. It is an understanding of the principles of government that is aimed at and their operations in the lives of men.

[10] *History and the Other Social Studies in the Junior High School* (Charles Scribners' Sons, 1926), p. 172.

Klapper[11] states the following three objectives in the teaching of civics:

1. To give an intelligent and sympathetic understanding of the interdependencies of the modern social life and the relation of individuals to the complex situation of living together.
2. To give knowledge of and desire to assume the specific duties of political citizenship. To develop civic sense and an interest in civic welfare.
3. To give an understanding of the structure and functions of government.

Measurement in the field of civics is limited to a few standardized tests. In general the measurement is as satisfactory as could be expected under the condition of change now existing in the field of social science.

The Brown-Woody Civics Test has been carefully developed. It has high reliability, .92. Its validity is as high as is usually attained in comparison with teachers' marks, being .65. The content is valid in that it is based on the analysis of content of nine widely used civics texts. No material is included that is not found in at least five of the nine texts. The manual of directions is excellent. End of the year norms are available for junior and senior high school. Percentile norms are given. Two forms are available.

The test is made up of three parts.[12]

Test I is designed to determine the extent of the pupil's mastery of the vocabulary encountered in the literature of civics. It consists of forty words selected at random from the list of words common to at least five of the nine textbooks, which were considered essential to an intelligent reading of the subject matter of the civics textbooks. This test is an "alternative answer" test, in which the pupil must select one of four given words or expressions which most nearly means the same as the indicated word in the vocabulary test. In building this test there was a definite attempt to incorporate words or ex-

[11] *The Teaching of History* (D. Appleton and Co., 1926), p. 301.
[12] Published by the World Book Company, 1926.

pressions which were closely enough related to the given word in meaning or in form to insure a careful discrimination in choosing the proper response.

Part I. Civic Vocabulary

Directions: Draw a line under the word, or group of words, in parentheses, which, in a civic sense, most nearly means the same as the first word in each line.

Begin here.

1. statute—(law, constitution, tradition, custom) 1
2. thrift—(stinginess, riches, greed, economy) 2
3. coöperation—(unselfishness, working-together, harmony, without friction) 3
4. urban—(pertaining to travel, pertaining to country, pertaining to city, pertaining to street cars) 4
5. treason—(punishment, defeat, betrayal, defiance) 5
6. federal—(local, state, national, American) 6
7. wages—(pay, check, money, credit) 7
8. authority—(influence, control, action, commission) 8
9. occupation—(work, job, task, vocation) 9
10. jurisdiction—(enforcing law, granting a franchise, establishing a principle, having authority) 10

Test II is designed to measure the extent of the pupil's information in the special field of civics. It consists of eighty "yes or no" questions common to at least five of the nine textbooks. The questions are distributed over local, state, and federal government in approximately the same proportion as revealed in the analysis of subject matter in the textbooks. By including as many as eighty questions this test is of sufficient length to give reliable measurement.

Part II. Civic Information

Directions: Draw a line under the right answer to each question.

Begin here.

1. Is the United States a democracy? Yes No 1
2. Is the Constitution of the United States the highest law of the land? Yes No 2
3. May any adult become a candidate for office, local or national? Yes No 3
4. As a general rule does ignorance of the law excuse its violation? Yes No 4
5. Is the President of the United States the executive officer of the nation? Yes No 5
6. Is a bill the same thing as a law? Yes No 6

7. Does the Secretary of State have charge of military affairs? Yes No 7
8. May a person accused of crime be tried secretly? Yes No 8
9. Is New York the nation's capital? Yes No 9
10. Must one be 21 years of age to vote at a presidential election? Yes No 10

Test III is designed to measure how well the pupil can apply the principles of civics to practical situations. The responses to these situations demand the application of what may be termed "civic thinking," and at the same time it is possible through them to determine whether the pupil possesses proper civic attitudes, ideals, and appreciations.

Part III. Civic Thinking

Section 1

Directions: Imagine that you are a voter. There are two candidates seeking the office of mayor of your city. The qualities and characteristics of each are listed below. We shall call one Mr. A and the other Mr. B. You are to select the candidate who is, in your opinion, best fitted to perform the duties of that office. Read carefully the characteristics and qualities of each before you make your decision. Place an X in the square following the name of the candidate for whom you would vote.

Begin here.

Mr. A ☐	Mr. B ☐
Legally qualified for office	Legally qualified for office
Relative of Governor of the State	Experienced as a lawyer
Ex-senator	Previously an alderman of his ward
University graduate	Junior College graduate
Member of Board of Commerce	Member of Board of Commerce
President of an industrial corporation	Successful business man in the community
Member of several fraternal orders	Interested in city's welfare
"Friend of the laboring man"	Far-sighted
Wealthy	Conscientious
Critical of city management	Honest in dealing
Well known in social circles	Prominent in business circles
Church member	Sterling character
Fluent speaker	"Man of few words"
Conservative	Progressive
Friendly	Intellectual
Efficient politician	Organizer

After you have selected one of the above candidates, place a check mark (✓) before the qualities or characteristics that determined your choice.

7. Mr. L lives on a little farm just outside the limits of the city of X in the state of R. Mr. L wakes up one morning and finds that thieves have taken grain, chickens, and other possessions from his premises during the night. Mr. L wishes to bring the thieves to justice. To whom should he report his losses?

1. The chief of police of the city of X.
2. The chairman of the Board of Public Safety of the city of X.
3. The State Department of Justice in the state of R.
4. The supervisor of his township.
5. The county sheriff.

This test measures to a marked degree the objectives of civics teaching quoted from Knowlton.

Much effort is now being made to develop in the pupils while they are in school the proper citizenship traits and attitudes that are so much needed in adult life. The first step in doing this is to make the pupil a good school citizen. This requires two things—an analysis of what these traits and attitudes are, and some method that will tend to emphasize their importance and make possible practice in them.

An attempt has been made to develop a quantitative measure of the progress of pupils in forming habits and attitudes of good citizenship. These habits and attitudes refer primarily to the school citizenship of the child. The Chassell-Upton Citizenship Scales [13] were developed principally to aid teachers in the training of their pupils in certain particular habits and attitudes. They are valuable in that they point out the particular habits and attitudes in which the child is weak or deficient. In addition, they give practice in the development of initiative and power of self-direction by children while they are in school. There are eight short scales or charts—A to H. Chart C is given in Figure 17.

For full directions for using and scoring the scales, see the *Teachers College Record* for January, 1922. [A part of these directions are given on pages 261 and 262.]

[13] Published by Teachers College, Columbia University, 1922.

Score in Points 195 Average Score in Per Cent
Score in Per Cent 52 on Scales C and.....
Name John Jones Age....yrs.....mos. Sex..... Grade....
School XYZ Pupil marked by..... Date.....

3 × 8 = 24	0	1	2	3	Puts on or removes wraps quickly.
	0	1	2	3	Turns to the right in passing people.
	0	1	2	3	Keeps regulations relating to the recess period.
	0	1	2	3	Accepts no "dares" of a foolhardy nature.
	0	1	2	3	Does not cry or complain over trivial things.
5 × 16 = 80	0	1	2	3	Allows pupils in front of him or older persons to pass through doorways or into the elevator first.
	0	1	2	3	Reports symptoms of illness promptly.
	0	1	2	3	Enjoys a joke, even at his own expense.
	0	1	2	3	Reads assigned work whether he is to be questioned on it or not.
	0	1	2	3	Does not expect special favors or privileges.
	0	1	2	3	Does not intrude on another person's privacy or time.
	0	1	2	3	Takes pride in the appearance of school property, doing his part to keep the cloak-room in order; picking up paper from floor and stairs; and taking pains not to throw orange-peel or other refuse on street or playground, or to deface desks, books, pictures, walls, woodwork, etc.
	0	1	2	3	Honors those who have rendered distinguished service.
	0	1	2	3	Seeks information by asking questions, by observation, and by other methods.
	0	1	2	3	Is agreeable when he can not have his own way.
	0	1	2	3	Finds possible solutions for problems.
7 × 13 = 91	0	1	2	3	Is reverent in attitude during prayer, or at least does not disturb the devotions of others.
	0	1	2	3	Assists heartily in measures to prevent the spread of disease and sickness, such as taking care in case of colds.
	0	1	2	3	Endeavors to restore lost property to the rightful owner.
	0	1	2	3	Concentrates upon the task at hand.
	0	1	2	3	Forms worthwhile purposes and strives to carry them out.
	0	1	2	3	Shows a spirit of helpfulness and service to others, whether in work or play.
	0	1	2	3	Works for his team rather than for himself.
	0	1	2	3	Is democratic.

FIG. 17.—Chassell-Upton Citizenship Scales—Short Scale C.

The following directions, applying to Chart C, explain how to use these charts.

The numbers 0, 1, 2, and 3 appearing in the columns at the left of each item on the scale, see Figure 17, are to be interpreted as follows:

0 indicates almost complete absence of the habit or attitude in question;
1 indicates below average attainment;
2 indicates above average attainment;
3 indicates practically perfect attainment.

In the case of each item underline the number which represents the child's present degree of attainment in the given habit or attitude. After marking the twenty-four items in this way, proceed further as follows:

(1) Add the underlined numbers opposite the first five items (separated from the remainder of the scale by a black line extending across the page) and enter this sum in the space provided after the multiplier 3 in the margin at the left. Perform the multiplication indicated, and enter the result in the space provided.

In the specimen marking of Scale C, given in Figure 17, the sum of the five underlined items in the first part is 8, which is to be multiplied by 3, thus giving 24 as the score on the first part of Scale C.

(2) Add the underlined numbers opposite the eleven items in the second part of the scale, enter this sum in the space after the multiplier 5, perform the multiplication indicated, and enter the result.

In the specimen marking of Scale C, the sum of the eleven underlined items in the second part is 16, which is to be multiplied by 5, thus giving 80 as the score on the second part of Scale C.

(3) Add the underlined numbers opposite the eight items in the third part of the scale, enter this sum in the space after the multiplier 7, perform the multiplication indicated, and enter the result.

In the specimen marking of Scale C the sum of eight underlined items in the third part is 13, which is to be multiplied by 7, thus giving 91 as the score on the third part of Scale C.

(4) Add the scores on the three parts of the scale, and enter the total score at the top of the page, after the words Score in Points. The maximum score, possible only in case an individual should be considered of sufficiently perfect attainment for a credit of 3 to be assigned to each item, is 378 points.

In the specimen marking of Scale C the addition of 91 + 80 + 24 gives 195 as the total score, or the score in points.

(5) Find the value of this score in points expressed as a per cent of 378 (the maximum score possible), and enter this figure after the words Score in Per Cent, in the space provided at the top of the page.

The total score of 195 on Scale C is equal to 52 per cent of 378, the maximum score possible. Thus 52 is entered as the score in per cent.

While the judgments to be made in connection with the use of these scales are somewhat subjective and the norm in a given habit or attitude is not very well known, nevertheless these scales may well serve several very important functions:

1. Their use places the emphasis on development of habits and attitudes, which is much neglected in the school curriculum.
2. The list of habits and attitudes included in the scale is a valuable list for pupils and teachers to have before them.
3. The use of the scale gives an objective measure of desirable phases of school work. The use of the scale will tend to point out specific needs of particular children.
4. The use of the scale by the children will do much to develop self-direction and self-measurement, which development is highly desirable.

TESTS IN HISTORY AND CIVICS

SOCIAL SCIENCE

Kepner Background Test in Social Science. By Tyler Kepner. Published: 1924. Purpose: to measure factual background for social science study. Range: grades 9-12. Time to give: 40 to 45 minutes. Time to score: 4 minutes. Not standardized. Publisher: Harvard University Press.
Price: $1.25 per lot of 25, complete; specimen set, 25 cents.

CIVICS

Brown-Woody Civics Test. By Arold W. Brown and Clifford Woody. Published: 1926. Purpose: to measure achievement in civics. Range: grades 7-12. Time to give: 35 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.30 per package of 25 examination booklets; specimen set, 15 cents.

Burton Civics Test. By W. H. Burton and Virginia N. Burton. Published: 1928. Purpose: to measure information in civics. Range: grades 5-9. Time to give: 20 to 25 minutes. Time to score: 3 to 4 minutes. Standardized.
Publisher: World Book Co.
Price: no information.

Odell Scales for Rating Pupils' Answers to Nine Types of Thought Questions in Civics. By Charles W. Odell. Published: 1927. Purpose: to measure achievement in civics. Range: junior and senior high school. Time to give: no time limit. Time to score: short. Standardized tentatively.
Publisher: Bureau of Educational Research, University of Illinois.
Price: no information.

Upton-Chassell Short Scales for Measuring Habits of Good Citizenship. By L. M. Chassell, Siegried M. Upton, and C. F. Chassell. Published: 1922. Purpose: to measure pupils' progress in good citizenship. Range: grades 1-8. Time to give: no information. Time to score: short. Standardized.
Publisher: Teachers College, Columbia University.
Price: 50 cents per 100. Copy of *Teachers College Record*, January, 1922, 40 cents.

AMERICAN HISTORY

Barr's Diagnostic Tests in American History. By A. S. Barr. Published: 1918. Purpose: to test pupils' general historical information. Range: grades 8-12. Time to give: 30 minutes. Time to score: 2 to 3 minutes. Standardized.
Publisher: Public School Publishing Co.
Price: $4 per 100, Series 2A or 2B; specimen set, 15 cents.

Columbia Research Bureau American History Test. By Harry J. Carman, Thomas N. Barrows, and Ben D. Wood. Published: 1926. Purpose: to test general achievement in Ameri-

can history. Range: grades 9 through college. Time to give: 125 minutes. Time to score: 5 minutes. Standardized.
Publisher: World Book Co.
Price: $1.60 per package of 25.

Ely-King Tests in American History. By Lena A. Ely and Edith King. Published: 1926. Purpose: to measure achievement in American history and diagnose difficulties. Range: grade 8. Time to give: 15 to 20 minutes. Time to score: 5 minutes. Standardized.
Publisher: Research Publishing Co.
Price: 75 cents per package of 25, any test; $2 per 100, $17.50 per 1,000; specimen set, complete, 40 cents.

The Gregory American History Tests. Test III. By C. A. Gregory. Published: 1923. Purpose: to measure pupils' information in American history. Range: grades 8-12. Time to give: 35 minutes. Time to score: 3 minutes. Standardized.
Publisher: C. A. Gregory, University of Cincinnati.
Price: $3.50 per 100 tests, complete; specimen set, 10 cents, postpaid.

The Hahn History Scales. By H. H. Hahn. Published: 1920; revised, 1922. Purpose: to measure the pupil's information, comprehension, and judgment in American history. Range: grades 7 and 8. Time to give: no time limit. Time to score: varies. Standardized.
Publisher: Public School Publishing Co.
Price: 20 cents each; 3 or more copies, 16 cents each; extra record sheets, 5 cents each.

Odell Scales for Rating Pupils' Answers to Nine Types of Thought Questions in American History. By Charles W. Odell. Published: 1927. Purpose: to measure achievement in American history. Range: junior and senior high school. Time to give: no time limit. Time to score: short. Standardized tentatively.
Publisher: Bureau of Educational Research, University of Illinois.
Price: no information.

Pressey-Richards Test for the Understanding of American History. By L. W. Pressey and R. C. Richards. Published: 1922. Purpose: to test information and judgment. Range: grades 6-12. Time to give: 30 minutes. Time to score: 2 minutes. Standardized.
Publisher: Public School Publishing Co.
Price: $2 per 100 copies; specimen set, 10 cents.

Van Wagenen American History Scales (Revised Edition). By M. J. Van Wagenen. Published: 1925. Purpose: to measure

achievement in American history. Range: grades 5-8: Scale 1, grades 5 and 6; Scale 2, grades 7 and 8. Time to give: no information. Time to score: no information. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: Information Scales, $2 per 100, each scale complete (indicate grade); Thought Scale R2, $2.25 per 100 complete; extra manual of directions, 25 cents; extra record sheets, 10 cents each.

EUROPEAN HISTORY

The Gregory-Owens Tests in Mediæval and Modern History. By C. A. Gregory and A. D. Owens. Published: 1926. Purpose: to test pupils' knowledge in medieval and modern history. Range: grade 11 through normal school. Time to give: 40 minutes. Time to score: 3 minutes. Standardized.
Published: C. A. Gregory, University of Cincinnati.
Price: $4 per package of 100. In quantities less than 100, 4½ cents per copy and 5 cents extra for score card and class record sheet; specimen set, either form, postpaid, 10 cents; both forms, 20 cents.

Vannest Diagnostic Test in Modern European History. By C. G. Vannest. Published: 1921. Purpose: to diagnose pupil difficulties in history. Range: secondary school. Time to give: 40 minutes. Time to score: 15 minutes. Standardized tentatively.
Publisher: Bureau of Coöperative Research, Indiana University.
Price: 1 to 10 copies, 10 cents each; 11 to 50 copies, 2 cents each; 51 to 100 copies, 1½ cents each; extra directions sheets, 2 cents each; extra scoring keys, 5 cents each.

BIBLIOGRAPHY

Barr, A. S. "Diagnostic Tests in American History," *Journal of Educational Research,* Vol. 3: pp. 74-75, January, 1922.

Barr, A. S. "Measurements in Civics," *Historical Outlook,* Vol. 17: pp. 24-26, January, 1926.

Brinckley, S. G. "Values of New-Type Examinations in the High School, with Special Reference to History," *Teachers College Contributions to Education,* No. 161, Columbia University, 1924.

Buckingham, B. R. "A Proposed Index of Efficiency in Teaching United States History," *Journal of Educational Research:* pp. 161-72, March, 1920.

Buckingham, B. R. "Correlation Between Ability to Think and Ability to Remember with Special Reference to United States History," *School and Society,* Vol. 5: pp. 443-49, April 14, 1917.

Dunn, A. W. *The Social Studies in Education,* Bulletin No. 28, Department of the Interior, Bureau of Education, Washington, D. C., 1916.

Elston, Bertha. "Improving the Teaching of History in the High School Through the Use of Standard Tests," *Historical Outlook,* Vol. 14: pp. 300-04, 1923.

Gibson, O. H. "Existing Standard Tests in History," *Historical Outlook,* Vol. 12: pp. 324-26, December, 1921.

Gold, M. S. "Testing Vocabulary in History," *Historical Outlook,* Vol. 17: pp. 285-91, October, 1926.

Hardy, R. E. "New Types of Tests in Social Science," *Historical Outlook,* Vol. 14: pp. 326-28, November, 1923.

Harlan, Charles L. "Educational Measurement in the Field of History," *Journal of Educational Research,* Vol. 2: pp. 849-54, December, 1920.

Kepner, P. T. "A Survey of the Test Movement in History," *Journal of Educational Research,* Vol. 7: pp. 309-25, April, 1923.

Odell, Charles W. "The Barr Diagnostic Tests in American History," *School and Society,* Vol. 6: pp. 501-03, October 28, 1922.

Rugg, E. U. "The Character and Value of Tests in History," *School Review,* Vol. 27: pp. 757-71, December, 1919.

Sackett, L. W. "A Scale in United States History," *Journal of Educational Psychology,* Vol. 10: pp. 345-49, September, 1919.

Sackett, L. W. "A Scale in Ancient History," *Journal of Educational Psychology,* Vol. 8: pp. 284-93, May, 1917.

Vannest, C. G. "Diagnostic Test in Modern European History in Eighth Annual Conference on Educational Measurements," *Bulletin of the Extension Division,* Indiana University, Vol. 6: No. 12, pp. 59-68, 1921.

Van Wagenen, M. J. "Revised Van Wagenen History Scales," *Teachers College Record,* Vol. 27: No. 2, p. 142, October, 1925.

Van Wagenen, M. J. "Some Implications of the Reversed V. W. History Scales," *Teachers College Record,* Vol. 27: pp. 42-48, October, 1925.

Wessel, H. M. "A Critical Study of the Results of the Barr Diagnostic Tests in American History," *Historical Outlook*, Vol. 17: pp. 230-42, May, 1926.

Wilson, H. E. "The Continuity Test in History Teaching," *School Review*, Vol. 34: pp. 679-84, November, 1926.

CHAPTER XI

Geography

Following are the objectives of geography teaching as found by Stallman,[1] who made a study of the literature on geography teaching, of courses of study, and of administrators' and teachers' opinions as to what are proper objectives in the teaching of this subject. These statements of aims represent a composite from all these sources.

1. To develop an understanding of the effect of man's physical environment upon his life.
2. To develop correct habits of thinking in the solution of social problems.
3. To develop an understanding of man's relation to man socially and economically.
4. To give a working knowledge of the information necessary in ordinary life concerning place, surface features, political divisions, natural phenomena, and of how to interpret such information in maps, charts, museums, etc.
5. To give the child an understanding of the influences of geographical conditions on political events.
6. To give a sympathetic understanding of the conditions and peoples of other countries.

From a glance at these six aims it is apparent that the present-day objectives of geography teaching emphasize the relation of man to his environment. At a time when much of the earth's surface had not been discovered and when the communications of men were so dependent upon their meager

[1] "An Investigation of the Aims of Elementary School Geography," unpublished thesis, Indiana University, 1926. See also H. H. Barrows and Edith P. Parker, "Elementary Geography: Objectives and Curriculum," *Elementary School Journal,* Vol. 25, March, 1925.

knowledge of the earth, the phase of geography that was of greatest importance was that concerning the shape of the earth, accurate knowledge of its surface, and the meteorological conditions of the newly discovered parts. Now that every land is more thickly settled and communication is easy, man's chief struggle is with the problems of living sympathetically and justly with his fellow men, and so geography has naturally become a social or human science rather than a natural or physical science. In a study of any country or region the answers to the questions *what kind of people live there, what do they do, why do they do it, why do they do it that way,* and *how are our lives related to them* will give an understanding that is largely social. The teaching of geography is turning much to this type of study and thus is becoming a social science.

Geography, like history, suffers from lack of agreement as to its aims, content, organization, and methods. While one teacher may emphasize the problem side of geography, another may emphasize the location and earth features as the most important. The relative importance of home geography—state and national geography—as compared to the geography of other parts of the world is not agreed upon by those who teach. The grade placement of the various units of geography further complicates an attempt to measure achievement in it. Some schools believe in and practice the two-cycle plan of organization of subject matter, while others believe in the one-cycle plan.

Measurement in geography is concerned largely with the measurements of memory of facts, of the skills of map location and interpretation, of technical vocabulary, of the association of many and various items of information, of the ability to interpret information in terms of relationship, of the ability to solve problems. The problems to be solved may be of the organized kind where all the facts are given and organized; or they may be of the creative type where all the facts and materials are provided, but the child is left to organize and

interpret these facts in relation to the problem. Standard tests in geography measure:

1. Information—such as place location, facts about places, physical phenomena, political divisions, products, industry, and trade.
2. Skill in map reading.
3. Association of cause and effect.
4. Solution of problems.

Geography tests measure to a large degree Objective 4 (p. 268) and to a lesser degree Objectives 1 and 6. These tests, so far, are deficient in that they have failed to measure adequately understanding of the associations or relation between man and his geographical environment; they have failed to measure the development of the attitude of "sympathetic understanding" toward other peoples. They have emphasized factual knowledge more than the interpretive side of geography. Practically no inventory tests in geography have been developed. It seems to be an assumption on the part of teachers as well as on the part of those who would measure progress in geography that a pupil does not know any geography until he studies formal geography. As a matter of fact, any test of geography applied to pupils who are just taking up the study will reveal the fact that they already have a considerable knowledge of that subject. There is much need of inventory tests. This is especially true where geography is taught by the cycle plan. Diagnostic testing in this subject has not been extensive enough to be of any great value in pointing out specific difficulties that will provide an adequate basis for remedial teaching.

The Hahn-Lackey Geography Scale was developed by an analysis of the content of the six most used textbooks. The scale is composed of 224 well chosen questions. The scale is divided into 25 steps, each of which has in it a list of questions of practically equal difficulty. The difficulty of the various steps is indicated by the number of questions missed by a large

number of children. The estimated difficulty of any given step is stated for different grades. By choosing properly, a teacher can use the questions included as an examination. Almost any teacher of geography will find the scale suggestive for examination purposes. The teacher merely selects the questions she desires to use and writes them on the board as in a regular examination. Table 29 shows three steps with their difficulty for each grade and typical questions from these steps.

To obtain the score, divide the total number of correct answers written by the number of questions in the test. A sheet of directions for scoring, with a statement of what to accept and what to reject, is available. The list of questions is differentiated into tests of memory of geographic facts, which are shown in light-face type, and tests of thinking ability, which are shown in black-face type. Certain of the questions are listed by number on the bottom of the scale as testing "knowledge of the meaning of technical terms and symbols," "knowledge of the map as a geographical tool," etc. These questions cover more of the objectives of geography teaching than is usually found in any single test. In a sense this device makes the test somewhat diagnostic. The test is cumbersome to score. One of the writers took the questions in Column K, attempted to turn them into new-type tests, and gave them to children of the sixth grade. The results obtained made it clear that this could be done with the scale in such a way as not to lose much of its value as a measuring instrument. The questions as they are stated in the scale need some revising, especially if used in the lower grades, since the method of statement is a little confusing to younger children.

The Buckingham-Stevenson Tests [2] are worthy of much consideration. The geography place tests in these tests cover the places of importance in (a) the United States, (b) the world. The content has been carefully developed. The scoring is objective. The test of Information and Problems on various

[2] Published by the Public School Publishing Company, 1926.

TABLE 29.—COLUMNS H, K, AND O OF THE HAHN-LACKEY GEOGRAPHY SCALE [3]

	Step H	*Step K*	*Step O*
Fourth Grade	2	8	27
Fifth Grade	6	16	42
Sixth Grade	8	21	50
Seventh Grade ...	16	34	66
Eighth Grade	16	34	66

123. Which way does your shadow point or extend just before sundown in mid-summer? In mid-winter?
152. How does Russia in Europe compare in size with the United States?
198. Give four marked illustrations of Man's skill in overcoming natural geographic barriers.
221. Between what two bodies of land is the Bering Sea? Dover? Skager Rak? Bab el Mandeb?
110. What is the "Fall Line," and why are a number of cities located on it?
144. Why is there a heavy rainfall in the Amazon valley?
157. What is the cause of rain?
177. The heat equator passes through New Mexico in July and through Argentina in January. Give the chief reason why.
200. Give the principal reason why the coastal cities of the U. S. have a more temperate climate than cities of the same latitude in the Great Central Plains.
120. What instruments are used to find out the heat and weight of the air?
140. Locate St. Louis and Cincinnati.
141. Draw a line showing the direction of the principal mountain system in each of four continents.
153. Name five important ports on the continent of Europe.
167. Name two religions of the people of Asia and give one nation belonging to each.
181. Name four industries based on animal products.
209. What and where are the following: Nova Scotia, Calcutta, Good Hope, St. Elias, Moscow, Bagdad?
210. What states would you cross or touch in going from Philadelphia to Denver?
213. Which is larger: Maine or your own state, and about how much larger?
222. Name four bodies of water you would sail through in going from Liverpool to Hong Kong by way of Suez Canal.
55. How are valleys made?
58. Where does the moon get its light?
66. Name two important crops raised in southern United States and tell why they grow there.
87. Name a mountain range and a river between Europe and Asia.
97. What direction is Australia from Asia and to whom does it belong?
128. Name two states wholly or partly in the Lake Plains of the United States.
14. Name a large lake entirely in the United States.
83. Why should you not expect Russia to have as many sailors as England?
113. Why are the Great Plains just east of the Rocky Mountains not good for farming?
151. In what part of Australia are most of the people found?
185. A doctor advises a patient to go to a region of light air and warm climate. Where would he go?

[3] Published by the Public School Publishing Company, 1922.

countries covers the most common information and vital problems concerning each country. These tests measure slightly Objective 3 and, to a greater extent, Objective 4. Part I consists of 50 items of information; Part II of 25 problems. The two examples which follow are taken from the test on the United States.

Part I. Information Test

Write the number of the correct or best answer to each question in the parentheses before the question.

( ) 1. Which state has the most beautiful mountain scenery?
1 2 3 4
Illinois, Michigan, Ohio, California.

( ) 2. Which state raises the most sugar cane?
1 2 3 4
Arkansas, Georgia, Massachusetts, Ohio.

( ) 3. Which state is most densely populated?
1 2 3 4
Alabama, Utah, Delaware, Iowa.

( ) 4. Which state produces the most oil (petroleum)?
1 2 3 4
California, Washington, Nevada, South Carolina.

( ) 5. Which is the greatest manufacturing state?
1 2 3 4
Illinois, Georgia, Montana, Arizona.

( ) 6. From which country does the United States import the most sugar?
1 2 3 4
England, Spain, Cuba, Australia.

Part II. Problem Test

Write in the parentheses the number of the correct or best answer.

12. Why are large tracts of land in the West which are naturally dry and isolated becoming prosperous agricultural districts?

1. Good roads make communication between towns possible.
( ) 2. Farming is a profitable occupation.
3. Modern machinery makes farming on a large scale profitable.
4. Irrigation has made the land productive.

14. Why can Cincinnati obtain coal cheaply from Pennsylvania?
1. Railroad connections are excellent.
( ) 2. Cincinnati uses a large amount of coal.
3. Pennsylvania has more coal than it can use.
4. Coal can be shipped by boats.

16. Why would many men in New England be without work if there was a failure of the cotton crop in the South?
1. Men from New England work in the cotton fields.
( ) 2. Less cotton would then be raised in New England.
3. Cotton cloth is very necessary to mankind.
4. The New England mills obtain their raw cotton from the South.

The Posey-Van Wagenen Geography Scales[4] are of two types. One type is made up of information questions which call for knowledge of plans, sources of products, methods of production, trade routes, etc. In each of the fourteen information tests there are 30 items, each of which is given a certain value. These scales are available for grades 5-6 and 7-8. The second type consists of thought questions, which require that the pupil draw inferences from the statements given. There are two thought scales, each containing 30 items of a given value. The thought scales are general in scope and available for grades 5-6 and 7-8. Some of the scales are general, others apply to particular countries. The following samples from the tests will show the type of measurement attempted.

[4] Published by the Public School Publishing Company, 1922.

INFORMATION S (GENERAL)

Division 2—Grades 7 and 8

Directions: Begin with question 1 and answer as many of them as you can. Answer them in order. If you cannot answer all the parts of any question, answer as many of the parts as you can. Start.

Group I (Average value 69.5)

1.(65)
What two bodies of water does the Panama Canal connect?

1. .....
2. .....

6.(70)
After each of these deserts write the name of the continent in which it is found.

(a) Arabia .....
(b) Sahara .....

12.(76)
How would you describe the surface of each of these states: as mainly mountainous, a plain, or a plateau?

(a) Kansas .....
(b) Illinois .....
(c) West Virginia .....
(d) California .....

THOUGHT S

Division 1—Grades 5 and 6

Directions: In each of the problems below you will find one or more short paragraphs giving you certain facts. Below the paragraphs are questions or statements telling you to do something.

This is not a test to see if you remember something which you have learned. Read the paragraphs carefully and from the facts given in them think out the answers to the problems for yourself. Begin with problem 1 and do as many as you

can. Do them in order. If you cannot do all the parts of any problem do as many of the parts as you can. Start.

Group III (Average value 79.5)

21.(75)

On account of the lack of rainfall no other vegetation than grass is produced in abundance in the steppes of Central Asia and in Arabia. As a consequence the people do not live a settled life as we do but roam around from one place to another.

On what two kinds of food—fruits, meat, cereals, dairy products or vegetables—would you expect them to live?

.....

30.(84)

As the population in any region increases in density more food is needed. This must be raised nearby or else imported.

(a) In densely populated regions would you expect the land to be cultivated extensively, intensively, or carelessly?

.....

(b) Is the fertile land in densely settled regions likely to be more valuable or less valuable than in sparsely settled regions?

.....

These tests measure to some degree objectives 1, 3, 4, and 6 given on page 268.

The authors of these tests have given the following suggestions concerning the ones that are best to use under various conditions.

WHAT SCALE TO USE

For measuring the general range of information at the beginning or end of the year or for measuring the progress made by the pupils as the result of a year's work, use division 1 of scales R, S, or T for the fifth and sixth grades, and division 2 of scales R, S, T, U, V, or W for the seventh and eighth grades.

To determine the range of information at the close of studying definite continental divisions of geography, use scales A and

B for the geography of the United States and the rest of North America,—division 1 for the fifth and sixth grades and division 2 for the seventh and eighth grades; scales F and G for the geography of Europe—division 2 for the seventh and eighth grades; and scales K and L for the geography of the remaining continents as a unit—division 2 for the seventh and eighth grades.

There will occasionally be found a sixth grade pupil who answers all the tasks in the third group of tasks, making it impossible to find his score. In such cases give division 2 of the same or a similar test. In the same way, occasionally a dull pupil of the eighth grade who fails to do anything with the easiest group of tasks in the scale used may need to be given the next easier division of the scales in order to determine his score.

Thought Scale R, division 2, is especially adapted for measuring the ability of seventh and eighth grade pupils to handle problems in geography, while Thought Scale S, division 1, is designed for similar use with pupils of the fifth and sixth grades.

The scoring of the test is somewhat intricate. The Teachers' Handbook indicates the various answers that are correct or those that are to be given one-half credit. Standards are available. The test covers a wide range of material. Just how representative of the instruction in geography the content of the test is will depend upon how geography is taught. The measurement of the ability to think about geography questions, which the Thought Scale attempts to measure, is somewhat complicated at the best, because what is a thought question to one may be a memory question to another. However, in this same test for grades 5 and 6 the difficulty of reading comprehension enters into some questions to such a degree that it may invalidate the measure of ability to think on geography problems. Controlled-answer type questions would make this test much more easily used without materially lowering its value.

The Gregory-Hagerty Geography Tests for grades 4, 5, and

6 attempt to stress "the subject matter actually taught." They are issued in duplicate forms and are therefore usable as a measure of progress. They are very objective and easy to score, and have a high reliability. Norms are available by grades. Smith and Wright[5] have described the tests as follows:

This test consists of five parts. Part 1 consists of miscellaneous geographical facts of the completion type; Part 2 is a locational geography and map study test; Part 3 consists of questions and statements of the completion type on zones, seasons, climate, and rainfall; Part 4, same type as Part 3, on plants, animals, and peoples; Part 5, same type as 3 and 4, on countries and their products. Forms A and B are available.

A well-developed test in a particular skill is the Forney Test in Map-Reading Abilities.[6] This test is based upon a map of Africa. The pupil is given the political and economic map of Africa to study, and is asked to answer twenty multiple response questions, which measure his ability to interpret the legend scale of miles, the colors, the marginal signs, and various other symbols found on a map. The test is diagnostic of a pupil's ability to read and interpret maps. Some of the questions on the map are given as examples.

Draw a line under the number, abbreviation, word, or words that correctly answer each question or complete each statement.

Sample question: The capital of Madagascar (H6) is
Tamatave (H6) Tananarivo (H6)

1. Boundaries of countries are indicated by 1
red lines purple lines blue lines
8. The surface of Morocco (C1) is largely 8
lowlands young, rugged mountains uplands
and plateaus

[5] Second Revision of the Bibliography on Educational Measurements, The Bureau of Coöperative Research, Indiana University, November, 1927.
[6] Published by Ginn and Company, 1926.

13. In what direction is Cairo (G1) from Durban (G7)? 13
N. W. E. S.

16. About how many miles by air line is Cairo (G1) from Tunis (E1)? 16
700 1400 2000 3000

20. The latitude of Durban (G7) is 20
30° North 30° East 30° South 30° West

So little has been done in the diagnosis of pupil difficulties in geography that not much can be said concerning remedial work. The use of such a test as the one of map reading just described offers opportunity to get at the specific lack of a desired skill. The knowledge of the location of places and trade routes generally agreed upon as important may be tested, and remedial work and drill can make these basic facts well known. This procedure is justifiable provided more time and attention be given to the problem phases of geography and to its social, economic, and industrial implications. A better understanding of how children learn various phases of geography will do much to develop the diagnosis and remedial work in this subject. There is much need in this field for drill material that provides repetition of the informational facts which should be well known and that gives practice in the interpretation of geographical information as well as practice in problem solving.

TESTS IN GEOGRAPHY

Buckingham-Stevenson Information-Problems Test in Geography. By B. R. Buckingham, P. R. Stevenson, Douglas C. Ridgley, and Julia M. Shipman. Published: 1923. Purpose: to test pupils' knowledge and reasoning power in geography. Range: grades 5-9. Time to give: 6 to 8 minutes for each test. Time to score: short. Standardized: no information.
Publisher: Public School Publishing Co.
Price: $2.00 per 100 copies, for each test, complete; specimen set, 15 cents.

Buckingham-Stevenson Place Geography Tests. By B. R. Buckingham and P. R. Stevenson. Published: 1926. Purpose: to measure pupils' ability to locate places. Range: grades 5-8.

Time to give: no time limit. Time to score: short. Standardized tentatively.
Publisher: Public School Publishing Co.
Price: 20 cents per booklet (3 forms each of World and United States) with 6 class record sheets; ordered in quantity, 12 cents each; without record sheets, single copies, 15 cents; extra class record sheets, 1 cent each.

Buckingham-Stevenson United States Geography Information and Problems Test. By B. R. Buckingham and P. R. Stevenson. Published: 1923. Purpose: to measure information in geography. Range: grades 6-9. Time to give: 14 minutes. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: $2 per 100 copies.

Forney Test in Map-Reading Abilities. By Earl B. Forney. Published: 1926. Purpose: to determine the abilities of pupils to read and interpret maps. Range: grades 6-8. Time to give: no time limit. Time to score: short. Standardized tentatively.
Publisher: Ginn and Co.
Price: no information.

The Gregory-Hagerty Geography Tests. By C. A. Gregory and Mary E. Hagerty. Published: 1926. Purpose: to measure pupil's knowledge in geography. Range: grades 4-6. Time to give: 30 to 35 minutes. Time to score: 3 minutes. Standardized.
Publisher: C. A. Gregory, University of Cincinnati.
Price: $4 per package of 100. In quantities less than 100, 4 cents per copy and 5 cents extra for class record sheet and score card; specimen set, either form, 10 cents; both forms, 20 cents.

Gregory-Spencer Geography Test. By C. A. Gregory and Peter L. Spencer. Published: 1923. Purpose: to measure pupils' achievement in geography. Range: grades 6-8. Time to give: 35 minutes. Time to score: 3 minutes. Standardized.
Publisher: Bureau of Educational Research, University of Oregon.
Price: $4 per 100, each form, complete; specimen set, 10 cents, each form.

Hahn-Lackey Geography Scale. By H. H. Hahn and E. E. Lackey. Published: 1918; revised, 1922. Purpose: to test children in the mastery of the subject-matter of geography. Range: grades 4-8. Time to give: varies. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: 17 cents per copy.

Posey-Van Wagenen Geography Scales. By C. J. Posey and M. J. Van Wagenen. Published: 1922. Purpose: to measure pupils' ability in geography. Range: Division 1, grades 5 and 6; Division 2, grades 7 and 8. Time to give: 40 minutes. Time to score: 2 minutes. Standardized.
Publisher: Public School Publishing Co.
Price: $1.50 per 100, each division; Teacher's Handbook, 20 cents; specimen set, 30 cents.

Russell-Harr Diagnostic Geography Tests. By Russell and Harr. Published: 1926. Purpose: to determine the ability of pupils in the field of geography, and to compare results with results in other schools. Range: geography students. Time to give: no time limit. Time to score: short. Standardized: no information.
Publisher: McKnight and McKnight.
Price: 1 cent each for any number less than 100, plus postage. $1, less 20 per cent, plus postage, for 100 or more of one test or assorted tests, in one order. $10, less 30 per cent, plus postage, for 1,000 or more of one test or assorted tests in one order. Answer book with standardized scores, 20 cents.

BIBLIOGRAPHY

Barnes, Burton A. "Measuring the Effects of Supervision in Geography," *School and Society,* Vol. 10: pp. 61-70, July 19, 1919.

Barthelmess, Harriet M. "Geography Testing in Boston," *Journal of Educational Research,* Vol. 2: pp. 701-12, November, 1920.

Branom, M. E. *Measurement of Achievement in Geography.* The Macmillan Company, 1925.

Branom, M. E. "Objective Measurement of Problem Geography," *Journal of Geography,* Vol. 25: pp. 52-59, February, 1926.

Gregory, C. A. and Spencer, Peter L. "A Geography Test for the Sixth, Seventh, and Eighth Grades," *School and Society,* Vol. 15: pp. 452-56, April 22, 1922.

Ridgley, D. C. "1,200 Selected Place Names," *Journal of Geography,* Vol. 25: pp. 201-21, September, 1926.

Stevenson, P. R. "Minimum Essentials in Place Geography," *Journal of Educational Research,* Vol. 6: pp. 300-06, November, 1922.

Stevenson, P. R. "A Problem Test in Geography," *Journal of Educational Research,* Vol. 5: pp. 350-53, April, 1922.

Witham, E. C. "Standard Geography Tests," *American School Board Journal,* Vol. 71: pp. 51-52, November, 1925.

CHAPTER XII

Music and Art

Instruction in music as a phase of public school work has very greatly increased in importance during the past few years. In fact, an attempt is now being made to universalize music education. The discussion group on music education in the Department of Superintendence passed resolutions at the 1927 meeting and at the 1928 meeting which represent the attitude of educators toward a broad program of music education.[1] However, at present not more than five per cent of the total allotment of time on the school program is given to music.

In the upper grades and in the junior and senior high school, much time and effort is given to the development of a small per cent of the total number who can sing in the chorus or play in the orchestra or band. This development has come about because of a desire to recognize and to provide for individual differences and special ability, and also because such groups promote school and community morale and enjoyment. In many high schools also, elective courses in music are given, including history of music, harmony, music appreciation, and glee club.

The objectives of music teaching are concerned primarily with developing, on the one hand, certain skills and abilities in music: reading music, singing, playing instruments; and, on the other hand, appreciation and enjoyment. These two objectives are in no way opposed to each other and are usually concomitant. Music instruction was formerly largely confined to "note reading," and it is generally conceded that after many

[1] *Music Supervisors Journal*, May, 1928, p. 5.

years of such instruction, pupils did not show evidence of any very marked development of an appreciation of music. Educators have only recently discerned that enjoyment and appreciation of music are not necessarily confined to those pupils who by native capacity and training have acquired ability to read music and to sing or play, and have now begun to investigate the methods of teaching music in the public schools in a more scientific manner.

The Art of Music and the graphic and plastic arts are far behind other academic subjects in the attempt to apply the scientific method to their problem of subject matter, theory, and pedagogy. This is due to two principal causes: first, that the typical artistic mind is seldom interested in the scientific aspect of the art; and second, that these arts have not yet been thoroughly stabilized and assigned their proper places in the public school curriculum. . . . Because of the extravagant claims made by enthusiasts for music education and their indiscriminate attempts to force a flat universal music requirement upon every child in the public schools regardless of his innate capacities or abilities to profit by such education, the wholesome growth of music in the public schools has been seriously impeded.[2]

Educators have realized that the aims and objectives of school music should be clarified and some accurate means devised by which attainment could be measured. As a result, a field of scientific investigation has been developed which can be made productive of far-reaching results. This will depend on whether the results of measuring are utilized (1) to guide the individual pupil, and (2) to improve instruction to the end that the spirit and mechanics of music education may serve the all-round musical development of the pupil.

Measurement in music has been primarily of two types, measurement of musical capacity and measurement of musical accomplishment. The measure of musical capacity that has

[2] *Third Yearbook,* Department of Superintendence (February, 1925), p. 354.

been outstanding in its contribution to the scientific measure of a special ability is the Seashore Tests of Musical Ability.[3] These tests measure certain basic abilities or capacities that underlie talent in music. There are six tests developed on phonograph records in such a way that a careful teacher can easily administer them according to directions with the use of a phonograph. The six tests are tests of (1) Sense of Pitch, (2) Sense of Intensity, (3) Sense of Time, (4) Sense of Consonance, (5) Sense of Memory, (6) Sense of Rhythm.

In the first test, which is a test of pitch discrimination, after the pupil is given a blank on which he can record his judgments and the phonograph record is put on the machine, the pupil is instructed as follows:

> You will hear two tones which differ in pitch. You are to judge whether the second is higher or lower than the first. If the second is higher, record H; if lower, record L.

Following this instruction the preliminary test is given. In this test one hundred pairs of tones are sounded by playing the record on a phonograph. At first the pairs of tones vary enough in pitch, or number of vibrations, so that almost any pupil can distinguish between them. The difference in pitch between the tones in a pair is less with each succeeding pair until the tones differ only one-half vibration in pitch. This difference is not distinguishable by the person of ordinary pitch discrimination. Keys are available for scoring the answers of the pupils. The score in this test, as well as in all others of this series of Seashore tests, is the per cent of the number right. These scores can be turned into percentile scores by use of the percentile chart in the Manual of Directions.

The results on this test represent a measure of sensitiveness of the ear to pitch differences. According to Seashore, this ability or capacity cannot be appreciably improved. However,

[3] Published by the Columbia Phonograph Company, Inc., and based on C. E. Seashore's, *The Psychology of Musical Talent,* Silver, Burdett and Company, 1919.

the validity of this test as measured by correlation with the teachers' estimate of ability to sing is not high.

Test 2 is a test of capacity in discrimination of intensity or loudness of sound. The instructions to pupils are:

You will hear two tones which differ in loudness or strength. You are to judge whether the second is weaker or stronger than the first. If the second is stronger, record S; if the second is weaker, record W.

At first the distinction is easily made. The difference in intensity is made less during the test until it is barely distinguishable by those of greatest capacity in this innate ability. Since intensity of sound discrimination is dependent primarily on the delicacy of the structure of the ear, it would not be affected by training. While the predictive value of this test has not been proved, certainly insensitivity to sound below a certain critical point would make one unfit for development in music.

Test 3 is a test of the sense of time. The instructions to the pupils are:

You will hear three clicks marking off two intervals of time. If the second interval (that is, the time between the second and the third click) is longer than the first interval, record L; if it is shorter, record S.

The variations in difference between the two intervals of time are gradually made less. It is not assumed that the presence of this trait or capacity will guarantee ability to sing or play in time; "but defective sense of time must result in corresponding defect in marking of time in any form."

Test 4 is the test of the sense of consonance. The instructions to the pupils are:

You will hear two combinations of two tones each; one combination is better or worse than the other in consonance (harmony). A good combination is one in which the two tones are smooth, and blend, tending to fuse together into one. A bad

combination is just the opposite. If the second combination is better, record B; if worse, W.

"This calls for a judgment on blending, smoothness, and fusion, apart from the feelings of like or dislike, and apart from theory or feeling of musical value. Blending, smoothness, and fusion should be explained fully, and may be illustrated on the piano before the preliminary practice."

Test 5 is the test of tonal memory. The instructions to the pupils are:

In each trial you will hear a series of tones played twice. In the second playing, one note is changed. You are to record, by number, which one was changed. In listening count mentally; for example, 1, 2, in the first playing, and then likewise in the second playing, so that you may identify the one that was changed without error.

"There should be preliminary drill for each span of the five degrees of difficulty, with emphasis on the silent counting."

Test 6 is a test of sense rhythm. The instructions to the pupils are:

You will hear in rapid succession two rhythmic patterns; the second is either the same as the first or different. Listen and record either S for same or D for different.

The Seashore tests, since they test for fundamental capacities and not for achievement, do not necessarily correlate highly with some measures of musical achievement or with teachers' marks or judgments of achievement in music. (See the correlations with the Mosher Group-Test in Table 30.)

These correlations are low. However, the correlations shown here of a test of capacity or ability to do are not so much lower than the correlations sometimes found between intelligence test scores and achievement test scores. Since the inter-correlation between the various Seashore Tests is relatively low, it is fair to assume that a correlation between a composite of all of the five Seashore tests used and the Mosher test would give a cor-

TABLE 30.—CORRELATION OF MOSHER GROUP-TEST[4] WITH SEASHORE TESTS

	Time	*Consonance*	*Pitch*	*Memory*	*Intensity*
Coefficient of correlation...	.3581	.2912	.4391	.4386	.4858
Number of cases.....	459	434	460	430	460

relation coefficient as high as ordinarily found between intelligence test and achievement test scores given to the same pupils. Further evidence is needed along this line. The fact that a pupil has special capacity for music does not insure either his interest or his industry in music. The correlation between capacity and achievement is sometimes reduced because of learning difficulties or wrong methods of teaching. McConathy,[5] in a report on the use of the Seashore test as a phase of music work in certain schools, says that there is very definitely "a positive correlation between the findings of our fifth grade survey and the subsequent musical development of the child."

Seashore suggests that these tests be given in the fifth and eighth grades largely for purposes of guidance. Their value if used with some test of achievement in classification of pupils seems apparent. Further experimentation along these lines is needed.

There are several good tests of Musical Achievement. The Kwalwasser-Ruch Test of Musical Accomplishment[6] is based on the aims, materials, and attainments as set out by the Music Supervisor's National Conference and as found in the most representative courses of study. The test is made up of ten

[4] Raymond M. Mosher, "The Group Method of Measurement of Sight-Singing," Bureau of Publications, Teachers College, 1925, p. 45.

[5] "Public School Music," *Journal of Educational Research*, Vol. 14: No. 4, November, 1926.

[6] Published by the Bureau of Educational Research and Service, University of Iowa, 1924.

parts. Table 31 gives the maximum score and time for each of these parts.

TABLE 31.—THE PARTS, MAXIMUM SCORE, AND TIME LIMITS OF THE KWALWASSER-RUCH TEST OF MUSICAL ACCOMPLISHMENT

Test	*Maximum Score*	*Time Limit in Minutes*
1. Knowledge of musical symbols and terms..	25	5
2. Recognition of syllable names.....	25	4
3. Detection of pitch errors in a familiar melody	25	3
4. Detection of time errors in a familiar melody	15	3
5. Recognition of pitch names.....	25	4
6. Knowledge of time signatures.....	20	3
7. Knowledge of key signatures.....	30	4
8. Knowledge of note values.....	15	3
9. Knowledge of rest values.....	15	3
10. Recognition of familiar melodies from notation	50	8
	245	40

Following are samples from Tests 1, 6, and 10.

TEST 1. KNOWLEDGE OF MUSICAL SYMBOLS AND TERMS

DIRECTIONS: Below are twenty-five questions about music. Five answers are given to each question. Read each question and then draw a line under the right answer. The sample is already marked as it should be.

SAMPLE: ♩ is called a sharp natural flat note rest

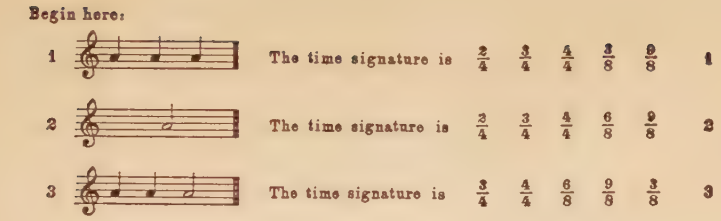
Begin here.

1	The first tone of the scale is	mi	re	do	fa	sol		1
2	♩	is called a	rest	natural	sharp	note	flat	2
3	The fifth tone of a scale is	do	fa	mi	sol	re		3
4	♯	is a	flat	note	natural	rest	sharp	4
5	♭	is a	sharp	flat	natural	note	rest	5
6		is a	slur	hold	rest	double-sharp	repeat-bar	6

TEST 6. KNOWLEDGE OF TIME SIGNATURES

DIRECTIONS: Below are ten full measures. At the right of each are five time signatures. You are to draw a line under the correct time signature for each measure. The sample is marked as it should be.

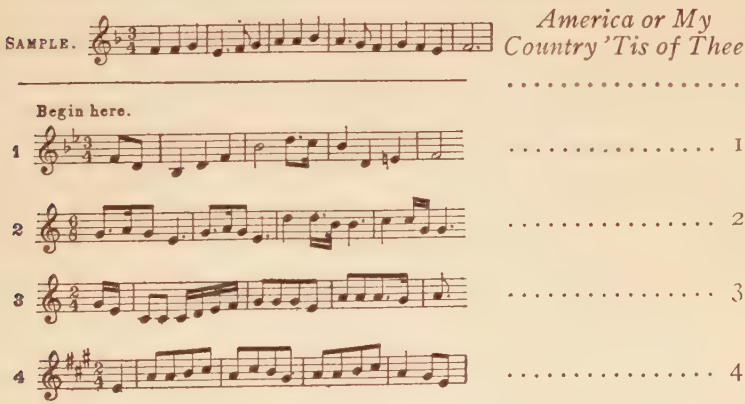
SAMPLE. The time signature is $\frac{2}{4}$ $\frac{3}{4}$ $\frac{4}{4}$ $\frac{6}{8}$ $\frac{3}{8}$



Test 10. Recognition of Familiar Melodies from Notation

Directions: Below are phrases from ten songs that you know. Hum each line to yourself and then write the name of the song or the words of the phrase on the line at the right.

The sample is already marked as it should be.



Only one form of the test is available. The test is highly reliable. It is very objective and easily scored. Norms are available for grades 4-12 inclusive. Percentile norms are also given. The test is valid in that it is based on and parallels a standard outline of the music course of study. Its validity as measured by its correlation with a music teacher's estimate of the child's musical accomplishment, in 54 cases in the sixth grade, was .62.[7] Much of the test is made up of measures of

[7] Data in the possession of the authors.

the pupil's knowledge of the symbolism of music notations. The test should be a valuable aid in determining the progress that is being made by the pupil on these phases of music. As a measure of whether pupils are accomplishing as a whole the standard attainment set up, it gives valuable aid in standardizing the course of study. As an example, Test 1, consisting of twenty-five items, measures the pupil's knowledge of musical terms and symbols. A measure with this test of the sixth grades in five large cities distinguished for their work in music revealed that the mean score on this test was only 13, a knowledge of approximately half of the most common symbols and terms.[8] Knowledge of the phases of music measured by this test may not be necessary to the life of the child. If such knowledge is necessary and is expected to be gained from music teaching, then it is fair to test for this achievement. The results obtained from the application of this test would reveal a deplorable condition in many places. Music teaching has in the past been guided by adult ideas of what the child should attain rather than by scientific knowledge of what the child can attain. The use of such a test as the Kwalwasser-Ruch test will do much to overcome this tendency. Certain parts of the test offer much opportunity for specific diagnosis of pupil difficulty. A study of the achievement of a pupil or class on the various phases of the test should point out what emphasis is needed in instruction. The use of the test to distinguish the outstanding accomplishment of those pupils with special ability or ability outside of the school training is marked. In the fifth grade class to whom this test was given by the writers, three pupils scored 170, 176, and 180. This is above the 12th year norm. In the same class one pupil scored 46 and another 60. These are below the norm of the fourth grade. It seems as though pupils in music might need classification upon the basis of accomplishment as well as ability.

[8] See *Third Yearbook*, Department of Superintendence (February, 1925).

At times, musicians have defended the grouping of inferior and superior in a class for the purpose of teaching, on the ground that the presence of the superior stimulates the inferior to greater accomplishment. Even if this attitude were demonstrable it would still be objectionable and indefensible. The presence of the superior pupils only heightens the inferiority consciousness of the inferior and tends to discourage them from realizing the meager results which otherwise would result from their efforts. Both groups are penalized by such an association and neither one does what it is capable of doing under more favorable circumstances. Whether music supervisors and teachers realize it or not, this evil of heterogeneous grouping is retarding progress in public school music education.[9]

In the Mosher Sight Reading Test, Mosher has developed a test in sight reading covering:

1. Knowledge of musical symbols, marks of expression, and general music information.
2. Recognition of scales, chords, and intervals.
3. Knowledge of measure and note values. (Duration, not pitch.)
4. Ability to identify well-known melodies when read silently.
5. Ability to write tonal figures or patterns from hearing them played in monotone on piano.
6. Ability to write rhythmic patterns from hearing them played in monotone on piano.
7. Ability to write melodies from dictation.[10]

The Hillbrand Sight-Singing Test is an individual test patterned after the Gray Oral Reading Test. It is primarily a diagnostic test to point out specific difficulties of pupils. The test is given in a quiet place. The pupil is given a copy of the test booklet and told, "'I should like to have you sing some of these songs for me. . . .' 'I am going to play the keynote of this song. . . .' 'I want you to sing the music using the

[9] John Kwalwasser, *Tests and Measurements in Music*, C. C. Birchard and Co., 1927, p. 36.

[10] "Group Method of Measurement of Sight-Singing," Bureau of Publications, Teachers College, Columbia University, 1925.

words.'"[11] The teacher holds a like copy of the song and notes mistakes on her sheet as the pupil sings. Time is to be kept on six songs sung. The errors noted by the teacher may be classified as belonging to one of these groups:

Notes wrongly pitched	Errors in time
Transpositions	Extra notes
Flatting	Repetitions
Sharping	Hesitations
Omission of notes	

Notes wrongly pitched. Whenever a pupil pitches any single note wrongly it may be because the note was misread, or it may be that the pupil has not mastered the interval to the note from the preceding one. In either case the error is counted simply as a "note wrongly pitched."

Indicate such an error by drawing a short line under the note, as shown at A in the sample song,[12] where the pupil sang E instead of F sharp.

Transpositions. If two or more consecutive notes are transposed as a group, either into another key or into a different position on the scale, it is considered that the error was made in the transposition. For example, let us suppose that the pupil sang the first measure of the third line on the same three notes as the preceding measure and then proceeded to sing the next measure correctly. It is considered that an error was made at B, where a transposition was made to a different part of the scale (or to a different key). This [error] is indicated by the letter *t* written under the note on which the transposition was made. Also, the pupil is considered as having made an error in transposing back to the right key at C, since if he does not correct the first error by singing the measure again, it is assumed that he did not notice that the interval from A to G sharp was only a half tone. If the pupil had sung the measure again without the transposition, both errors are considered to be corrected and are not counted as errors but as a repetition.

If the second transposition is made as a return to the original key, the letter *r* may be written also if desired to distinguish

[11] Manual of Directions of Hillbrand Sight-Singing Test, World Book Company.

[12] For this and following illustrations, see Figure 18.

this from a transposition to still a different key or position on the scale, which would be indicated simply by another letter *t*. The transpositions at B and C each count as one error.

If the pupil has wrongly pitched two or more consecutive notes, though not as a transposition,—as, for example, singing the G sharp and E in the next to the last measure as A and G sharp, respectively,—each of these two notes counts as a note wrongly pitched rather than as a transposition.

Flatting. If a single note is sung flat—as, for example, the note at D—this is indicated by writing the letter *f* under the note. If, however, the pupil flats appreciably on a note such as the note E, and continues to sing flat thereafter, this is indicated by the letter *f* followed by a dash. The *f*-dash at F indicates that the pupil flatted still more at that point. Each instance of flatting such as at D, E, or F, is counted as an error.

Sharping. Sharping is scored the same as flatting. The *s*-dash at G indicates that the pupil raised his pitch a little at that time.

Omission of notes. The omission of a note is marked by encircling, as in the case of the note at H. Each note omitted counts as an error.

Errors in time. The pupil may sing a measure wrongly as to time, as, for example, singing a $\frac{3}{4}$ measure in $\frac{2}{4}$ or $\frac{4}{4}$ time. If the error was made by shortening a note or notes, indicate this by flags on the stems of the shortened notes or by joining the stems as shown at M, which represents the case of the second and third notes sung as eighth notes.

If the error was made by lengthening a note or notes, it may be indicated by placing a dot after the note or in any convenient manner.

Each measure sung wrongly as to time counts as one error.

The three remaining types of errors are to be used for diagnostic purposes and while indicated on the first page of the test, they are not recorded on the Class Record and do not enter into the total error score of the pupil. These types of errors are found to occur but rarely.

Extra notes. The insertion of an extra note is marked as shown at J. In this case the pupil inserted the extra note, C, and also added an extra word, the, to go with it.

Repetitions. The repetition of a note or group of notes is indicated as shown at K. In this case the pupil repeated the three notes of the measure.

Hesitations. These are marked as shown at L. In this case the pupil hesitated a few seconds before continuing the singing.

It is sometimes difficult to decide just how to designate an error,—for example, whether it is a case of sharping or of transposition. One can never know for certain whether a note is wrongly sung because it was misread or because the interval was not familiar to the pupil. However, the examiner should indicate the nature of each error as nearly as possible. In any case in which the examiner is unable at the time to decide how to mark an error, he should underline the part of the song not properly sung and decide, if possible, at the conclusion of the singing how the error should be marked. Be sure to put down some mark for each error. If not sure that an error was made, give the pupil the benefit of the doubt. If the pupil has a slight foreign accent, or throat trouble of any kind, distinguish carefully between results of these defects and real errors.

After the test has been given, count the number of errors of each type and record the number in the appropriate space on the first page of the test. In thus recording errors, no distinction need be made between errors of the type D and errors of the type E. That is, all instances of flatting are counted together, all instances of sharping are counted together, etc. Find the total number of errors as indicated, not counting extra notes, repetitions, or hesitations.

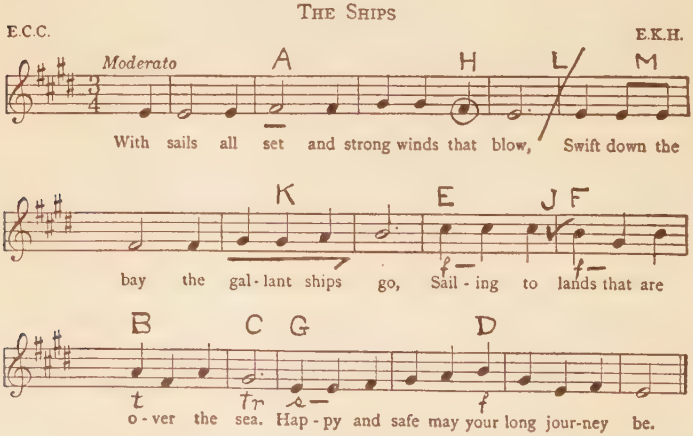


FIG. 18.—Song Illustrating Hillbrand Sight-Singing Test.

The measurement of Appreciation of music is difficult. Since appreciation is an outstanding objective in music teaching, attempts to measure accomplishment in it are worthy of note. A report by Trabue of Mohler's work in this line describes this attempt in some detail. The test was made up of phonograph records of sixteen different musical compositions. These were arranged into a scale on the basis of expert judgments.

The tests are arranged and administered in such a manner that a person who can detect small differences in the general merits of two selections will receive a high score, while one who can detect only the larger differences will receive a lower score. Three or four records are played in a group, one record immediately after the other. Each listener is asked, when a group of records has been played, to make a note indicating which record he considered "best," which "next best," and which "poorest." The differences in musical values of the records in the first group in each test are relatively large, while the differences in each succeeding group are smaller and smaller. The number of points allowed as a score for successfully rating each group of records is dependent upon the size of the errors in one's judgment, no credit at all being given if the person has rated the records in exactly the reverse order from that which is considered correct, and maximum credit being given when the correct order has been accurately indicated. A low score therefore indicates little or no success in judging the relative merits of these orchestral selections, while a high score indicates a greater degree of success.[13]

Two Courtis Standard Research Tests[14] in music attempt to measure (1) one's ability to associate life activities with music in which the rhythm predominates, (2) one's ability to recognize Mood from Melody. Samples from the tests follow.

[13] M. R. Trabue, "Scales for Measuring Judgment of Orchestral Music," *Journal of Educational Psychology* (December, 1923), pp. 543-60.

[14] Published by S. A. Courtis, 1807 East Grand Boulevard, Detroit, Michigan, 1922.

SERIES M

Test 1. Recognition of Characteristic Rhythms

For Grades Four to Twelve. (Experimental Edition)

Rhythm is one of the main elements in music. It has been defined as measured motion. Rhythmic motion also occurs in many of the activities of life. In this test you will be asked to judge from the music played, what life activity is represented.

5. When John returned at night he was tired and was glad to hear the sounds which told him that he was nearing his father's house. The music will tell you what John's father's occupation was. Underline the words which tell what the music says was John's father's business.

1. A blacksmith.
2. A clockmaker.
3. A preacher.
4. A miller.

Test 2. Recognition of Mood from Melody

Melody is the expression of a thought in music. In this test you will be asked to judge from the music played what John's thoughts were.

5. John's time was up now, so he took his pail and started for home. Listen to the selection and underline the words which best express how the music says John felt when his mother looked at what he had.
   1. He was *sorry* he had been cross about going.
   2. He was *glad* he had so many berries.
   3. He was *ashamed* that he had so few berries.
   4. He was *disappointed* that she said nothing.

ART

The Kline-Carey Measuring Scale for Free-Hand Drawing, Part 1—Representation is made up of four scales of representation. One scale is made up of drawings of a house and consists of twenty samples, the scale rating for the samples

varying from 0 to 95. Another scale consists of eighteen drawings of a rabbit, varying in value from 0 to 94. A third scale consists of sixteen drawings of a human figure in action, varying in value from 60 to 95. A fourth scale consists of the brush drawing of a tree, there being nineteen samples varying in value from 0 to 85. The samples that make up the scale were taken from pupils' drawings and were standardized according to the rating of experts. In the test the pupil is told to "draw a house, draw a rabbit, etc." The pupils' performance is then compared to the standard scale. The pupil's paper is marked according to the marking of the scale. The pupil himself can very well use the scale to score his work by moving his own drawing along the standard scale folder until a sample is found that appears to match. Under each of the standard scale samples is a legend which points out to the pupil the good points and defects of the drawing. It also indicates in what manner the particular sample is better than the sample before. For example, in the "house" scale under the sample marked "20" the following legend is given:

This drawing shows the corner view of a house. It stands upright but the gable ends should slant in the same direction. The door and chimney are well placed but too small, and the chimney is crooked. When you draw houses be sure to draw the upright lines first. Draw a house that stands straight, and that has straight doors, windows, and chimneys. Put it up by the side of this drawing and see if it is as good or better, and why.

Used in this way, the scale is somewhat diagnostic and instructive enough to be used successfully by the pupil. The scale has not been properly standardized but is in the process of standardization now. In general, art teachers do not favor an attempt to measure in the standard way the attempts of school children in drawing. However, certain values can be attained from the use of such a scale as this one.

TESTS IN MUSIC AND ART

MUSIC

Beach's Standardized Music Tests. By Frank A. Beach. Published: 1920. Purpose: to measure achievement in music. Range: grades 2-12. Time to give: 90 minutes. Time to score: 2 minutes. Standardized.
Publisher: Bureau of Educational Measurements and Standards, Kansas State Normal School.
Price: $2 net for 50 copies; specimen set and manual of instructions, 35 cents; specimen set containing manual, standard sheet, and test sheet, 50 cents, postpaid.

Hillbrand Sight-Singing Test. By E. K. Hillbrand. Published: 1924. Purpose: to obtain an objective measure of the ability of a pupil in the mechanics of oral sight reading in vocal music as an indication of progress made and the diagnosis of difficulties. Range: grades 4-6. Time to give: 10 minutes. Time to score: 5 minutes. Standardized.
Publisher: World Book Co.
Price: $1 per package of 25 examination booklets, complete; specimen set, 10 cents.

Hutchison Music Test. By Herbert E. Hutchison and L. W. Pressey. Published: 1924. Purpose: "to test ability in silent reading and recognition of musical scores from known songs and operas." Range: grades 7-12. Time to give: no time limit. Time to score: short. Standardized tentatively.
Publisher: Public School Publishing Co.
Price: 50 cents per package of 25; specimen set, 10 cents.

The Kwalwasser-Ruch Test of Musical Accomplishment. By J. Kwalwasser and G. M. Ruch. Published: 1924. Purpose: to measure objectively the progress of pupils in music. Range: grades 4-12. Time to give: 40 minutes. Time to score: 3 minutes. Standardized.
Publisher: Extension Division, University of Iowa.
Price: 6 cents per copy; $5 per 100 copies; specimen set, 10 cents, with manual, 16 cents.

Petry-Rasey Recognition of Characteristic Rhythms. By Harriet Petry and Marie Rasey. Published: 1922. Purpose: to measure pupils' ability to interpret life's activities and moods in terms of music. Range: grades 4-12. Time to give: 10 minutes. Time to score: short. Standardized.
Publisher: S. A. Courtis, 1807 East Grand Boulevard, Detroit, Michigan.
Price: $1.80 for set of 40 tests; $2.50 for set of 80 tests; specimen set, complete, 80 cents.

Seashore Measures of Musical Talent. (Based on the author's *Psychology of Musical Talent.*) By C. E. Seashore. Published: 1919. Five Columbia records. Purpose: to test fundamental musical capacity. Range: grades 5-8. Time to give: about 90 minutes. Time to score: 10 minutes. Standardized.
Publisher: Columbia Phonograph Company, Inc.
Price: $1.25 each record. Manual of instructions and interpretations free.

Torgerson-Fahnestock Music Test. By T. L. Torgerson and Earnest Fahnestock. Purpose: to measure attainment and progress of pupils in both the theoretical and practical aspects of music and to assist in diagnosing difficulties of pupils. Range: grades 4-8. Time to give: approximately 30 minutes. Time to score: short. Standardized tentatively.
Publisher: Public School Publishing Co.
Price: 75 cents per package of 25 copies each of Parts A and B, with direction booklet and class record sheet; specimen set, 15 cents.

DRAWING

Kline-Carey Measuring Scale for Free-Hand Drawing. By L. W. Kline and G. L. Carey. Published: 1922. Purpose: "to measure drawing as done in public elementary and high schools, and to measure the pupils' progress in that subject." Range: elementary and high schools. Standardized.
Publisher: The Johns Hopkins Press.
Price: 60 cents per copy.

Lewerenz Test in Fundamental Abilities of Visual Art. By A. S. Lewerenz. Published: 1927. Purpose: to measure progress in instruction in visual art and to select pupils for specialized art activities. Range: grades 3-12. Time to give: Parts I and II, 30 minutes each; Part III, 35 minutes. Time to score: short. Standardized.
Publisher: Research Publishing Co.
Price: Parts I and II, 65 cents per package of 25, $2 per 100, $17.50 per 1,000; Part III, $1 per package of 25, $3.50 per 100, $32.50 per 1,000; manual of directions, 35 cents extra; color chart (required for Part III), $4; extra vase forms, 5 cents each; extra vase form scoring keys, 5 cents each; specimen set, complete, $4.50; specimen set, without color chart, 50 cents.

Thorndike's Scale for the Merit of Drawings by Pupils 8 to 15 Years of Age. By E. L. Thorndike. Published: 1914. Pur-

pose: to measure ability in drawing. Range: grades 3-8, or 8-15 years of age. Time to score: short. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: 50 cents per copy.

Williams Scale for Judging Kindergarten Drawing. By Florence Williams. Published: 1922. Purpose: to measure ability to draw. Range: kindergarten and grade 1. Time to give: no time limit. Time to score: varies. Standardized.
Publisher: Department of Art Education, School of Education, University of Chicago.
Price: 5 cents each.

BIBLIOGRAPHY

Dykema, Peter. "Tests and Measurements in Music Education," *Proceedings of the National Association of Music Teachers, 1925.*

Gibson, K. "Experiment in Measuring Results of Fifth Grade Class Visits to an Art Museum," *School and Society,* Vol. 21: pp. 658-62, May 30, 1925.

Karwoski, T. F. and Christenson, E. O. "A Test for Art Appreciation," *Journal of Educational Psychology,* Vol. 17: pp. 187-94, March, 1926.

Kwalwasser, Jacob. *Tests and Measurements in Music,* C. C. Birchard and Company, 1927.

Mosher, Raymond M. "A Study of the Group Method of Measurement of Sight-Singing," *Contributions to Education No. 194,* Teachers College, Columbia University.

Schoen, Max. "Tests of Musical Feeling and Musical Understanding," *Journal of Comparative Psychology,* Vol. 5: pp. 31-51, February, 1925.

Schoen, Max. "Common Sense in Music Testing," *Proceedings of the National Association of Music Teachers,* 1925.

Schoen, Max. *The Effects of Music,* Harcourt, Brace & Company, 1927.

Schoen, Max. "Recent Literature on the Psychology of the Musician," *Psychological Bulletin,* Vol. 18: pp. 483-89, September, 1921.

Seashore, C. E. "Avocational Guidance in Music," *Journal of Applied Psychology,* Vol. 1: pp. 342-48, March, 1917.

Seashore, C. E. *The Psychology of Musical Talent,* Silver, Burdett and Company, New York, 1919.

Seashore, C. E. "Recent Progress in the Psychology of Musical Talent," *Music Teachers' National Conference Proceedings,* pp. 158-65, 1920.

Thorndike, E. L. "A Scale for Measuring Achievement in Drawing," *Teachers College Record,* Vol. 14: pp. 1-71, November, 1913.

Trabue, M. R. "Scales for Measuring Judgment of Orchestral Music," *Journal of Educational Psychology,* Vol. 14: pp. 545-61, December, 1923.

CHAPTER XIII

Science

The program of science in the public school is composed primarily of nature study in grades 5 and 6; general science in grades 7 and 8; biology or botany and zoölogy in grades 9 or 10; and physical geography, chemistry, or physics in grades 11 and 12. Various combinations of these subjects are made, and many of those subjects offered in large senior high schools are omitted in smaller schools. Agriculture, although taught somewhat as a science, is usually classed as a prevocational or vocational subject. There is relatively little of the total time of the elementary school given to science or nature study as a separate subject. Some science is taught, however, in connection with other subjects, especially geography. General science is pretty generally accepted as a constant in the junior high school program. The science courses of the senior high school are usually elective. Since the entrance requirements of most colleges include one year of a laboratory science, a very large number of high schools require this amount of science for graduation. The enrollment in science courses in the high school has greatly increased during the last twenty-five years.

The following incomplete list of objectives of the teaching of science is not applicable in its entirety to any one science, but represents some of the most commonly stated aims.

1. To give first-hand knowledge of the scientific facts, phenomena, principles, and laws of the various sciences as they apply to the lives of people.
2. To give ability to solve problems using these principles.
3. To develop scientific attitudes.
4. To develop scientific interests.

5. To develop the technique of laboratory procedure.
6. To do work preparatory to advanced work or direct vocational needs.
7. To train for worthy use of leisure time.
8. To adapt the child to his environment.

Some difficulties in curriculum making and in measuring the results of teaching in science have arisen due to differences in the purposes of the teaching of science in the high school. One value that need not be discussed is the propædeutic value. This is especially important to those pupils who are to attend technical schools or colleges. Certain specific vocational values are equally direct and important. A pupil in vocational agriculture is in need of certain specific phases of chemistry, physics, zoölogy, and botany. These may be, and usually are, taught in the agriculture course. In the objectives just stated, the emphasis is largely upon the values of utilization, consumption, understanding, and appreciation of science rather than upon technical knowledge and scientific production values. In the field of the natural sciences, as in so much of the present high school curriculum, the statement of objectives emphasizes social values. There still seems to be some confusion in the minds of teachers of science between the development-of-embryo-scientists purpose of teaching and the purpose that stresses the diffusion of knowledge of science for general consumption and application in the lives of people.

Measurement in science is largely divided into measurement in certain specific subjects such as botany, physics, and chemistry. The problems of measurement are of three kinds. First, there is a lack of agreement concerning what should be included in the course of study. Second, courses of study in the various sciences seldom include a statement of minimum essentials to be attained. Third, certain difficulties arise out of differences in emphasis given to different aims and objectives in the teaching of science. The emphasis in science instruction may be on the development of laboratory procedure and draw-

ing ability. If this is a legitimately expected outcome, it should be measured.[1] On the other hand, the emphasis may be almost wholly on the relation of the science studied to the life situations confronting the pupil, with very little thought given to the development of laboratory technique and drawing. Measurement in science has been deficient in two ways. In the first place, the emphasis has been too much on the informational and factual side of the subject and not enough on the measurement of interests and attitudes. Very little has been done, either, on measurement of an inventory type. Under the present system of science teaching much overlapping in subject matter is common. The high school chemistry teacher may assume that in the senior year pupils who have had general science do not know any chemistry, or she may assume that they know certain phases of practical chemistry that may be found in the general science course. The use of properly constructed inventory tests could do much to clarify this situation and to help teaching proceed on a basis of knowledge and not guesswork. Diagnostic tests and remedial work in the field of science have not progressed far enough; nevertheless the situation in this respect is better in science than in certain other subjects. The tests discussed in this chapter are divided into tests of general science, biology and botany, physics, and chemistry.

General Science

The Ruch-Popenoe General Science Test is published in two forms, *A* and *B*. This test purports to measure the achievement of pupils in general and elementary science. The content of the test is based on that found in the most commonly used textbooks. An analysis of the content of the test with reference to special science fields is shown in Table 32.

[1] The Department of Superintendence, *Sixth Yearbook* (1928), Chap. XVI, gives a good evaluation of laboratory and demonstrative methods in high school science.

TABLE 32.—SUBJECT-MATTER ANALYSIS OF THE RUCH-POPENOE GENERAL SCIENCE TEST IN PERCENTAGES

Biological science (botany, physiology, and zoölogy)	30%
Chemistry	12%
Physics and mechanical applications	38%
Earth science (agriculture, astronomy, geology, and physiography)	20%

The Ruch-Popenoe General Science Test is published in two forms designated as Form *A* and Form *B*. These are identical in construction and are of equal difficulty. Either form can be used alone, or both forms can be given and the results averaged. The use of both forms increases the reliability of the scores very considerably. One form can easily be given in a class period of forty-five minutes. The time limits have purposely been made generous enough to make the tests almost entirely power tests, and the slow pupil will not be penalized to any considerable extent. Increasing the time allowance has not been found to increase the scores significantly.[2]

Each of the equivalent forms of the test is made up of two parts.

Part I is composed of fifty items of general information concerning familiar elementary scientific facts, principles, concepts, terms, definitions, and applications.

An example of Part I is given:

Directions: Draw a line under the one word or phrase that makes the truest answer to the question, as shown in the sample.

Sample: An example of a liquid is
iron glass air <u>water</u> lead zinc oxygen

15. An organism which reproduces by means of spores is the
maple amœba bread mold sponge earthworm minnow cedar 15

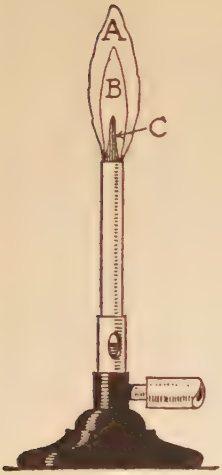
[2] G. M. Ruch and H. F. Popenoe, "Ruch-Popenoe General Science Test," World Book Company, Manual of Directions.

16. Sunlight can be broken up into the spectrum by means of a
telescope mirror burning glass lens prism microscope color-mixer 16
17. The act of transfer of pollen from anther to stigma is called
pollination reproduction fertilization transpiration mitosis adaptation filtration 17
18. Images are formed by the passage of light through a
prism helix ampere lens diaphragm mirage spectrum 18
19. The propelling mechanism of a machine is termed the
chassis piston motor differential carburetor governor fly-wheel 19
20. The separation of liquids and solids by evaporation and condensation is called
solution distillation diffusion fusion transpiration reduction pasteurization 20
21. A mirage is a kind of
swamp rainbow body of water optical illusion vision desert warfare 21
22. Decaying vegetable matter in the soil is termed
fungus nitrates mineral phosphates lichens potash humus 22
23. The smallest of the planets is
Mercury Venus Mars Jupiter Saturn Uranus Neptune 23

Part II measures the ability of the pupil to identify apparatus, organisms, structures and principles, and to apply principles of the science to the solution of simple problems.

Drawings and diagrams are used in this part of the test. An example is given:

Directions: On this page and the following pages there are 20 diagrams and several statements about each diagram. Part of each statement is left blank. You are to fill each blank so as to make a true statement. Make the statements as complete and definite as possible. In most cases you will need to write but *one word* in each blank. You will be given 25 minutes to do this part of the test.



In this diagram of the main parts of a flame:

a The burning zone is marked by the letter ..... *a*

b The region of unburned gas is marked ..... *b*

c The region made up of the products of combustion is marked..... *c*

d The chief products of combustion (burning) besides heat are:

1. .....
2. ..... *d*

The value of this test is largely in the help it should give the teacher in arriving at fairer final marks of achievement in her class. The results of this test, if it is given at the beginning of the term in connection with tests of general ability, should form the basis for classification. The reliability of the test is fairly high. The correlation between scores of pupils on Forms *A* and *B* is reported as .83. The test can well serve as a measure of progress if Form *A* is used at the beginning of the term and Form *B* at the close of the term. The use of the test in this fashion can be of much help in evaluating either methods of teaching or pupil activities. Grade and percentile norms are available. The test can be scored easily as it is highly objective. The time for giving is 40 minutes, 15 for Part I and 25 for Part II. The manual of directions and the scoring keys are well developed.

The Dvorak General Science Scales are available in three forms, Scales R-1, S-2, and T-2. S-2 and T-2 are equivalent forms. Scale R-1 is easier than the other two forms. The content of the test was taken from general science textbooks and questions submitted by experts. The resulting 600 ques-

tions were critically evaluated and then tried out experimentally. One hundred and eighty items selected from the original group of questions now make up the three scales. Each scale contains 60 items. The questions are of the single choice type. The author states that no item, regardless of its source, has been included which did not satisfy one or more of the general objectives of secondary education. The test is highly reliable and carefully standardized. Well-worked-out percentile norms are available.

The uses of this test are largely those of a survey test. While individual measurements with it are accurate and may well point out the pupil who is not making proper progress, yet the test is not diagnostic enough to give much aid in remedial teaching.

An attempt made by Curtis [3] to measure scientific attitude should be mentioned in this discussion. In attempting to get a measure of "Some Values Derived from Extensive Reading of General Science," a test in "Scientific Attitudes" was developed. This test, although not standardized, suggests certain possibilities for measurement. Part of the directions and part of the test are given.

On this sheet you will find a number of statements, problems, and short accounts about which you are to express your opinions. After some of the questions you will find several words or short sentences which apply to the question, and you are to make an X in front of the one which you think fits that particular question or statement best. For example:

A man picked up a piece of metal in the road and concluded that since he knew from the color of it that it was not gold, it must be tin. (*a*) Sound. (*b*) X Unsound. (You should write the X in front of the word "Unsound" because of course the conclusion is absurd.)

A child insisted that he had recently played with a real fairy. (*a*) It might be but isn't very likely to be true.

[3] "Some Values Derived from Extensive Reading of General Science," Teachers College, Columbia University, Contributions to Education, No. 163.

(*b*) It is doubtless true. (*c*) X It couldn't possibly be true. (*d*) It might be and probably is true. (You would mark the X in front of statement (*c*) which is the most appropriate of the four.)

A warden in a prison noticed that every man in the prison serving sentence on a theft or burglary charge had flat and blunt fingers. He concluded that flat and blunt fingers indicate a tendency to steal. (*a*) Sound. (*b*) X Unsound.

A certain aviator always carried a clover leaf when he flew, to give him good luck. One day he fell and was badly injured. He found that, without knowing it until after his fall, he had lost his clover leaf during the flight. He concluded that his fall was due to the loss of his good luck charm, the clover leaf. (*a*) The conclusion is sound. (*b*) X It is absurd. (*c*) It might be and probably is sound. (*d*) It might be but is unlikely to be sound.

Such a test might be developed and standardized so that it would be of much value in measuring an important aim in science teaching—"To develop scientific attitudes."

BIOLOGY

The Ruch-Cossmann Biology Test is available in two duplicate forms, *A* and *B*. The test was developed by the analysis of a large number of examination questions of good biology teachers from widely distributed parts of the country. The material was critically rated by good authorities. The test includes 112 items. The content is about evenly divided between botany and zoölogy. It takes 38 minutes of working time to take the test. There are five parts to the test. Tests I and II are illustrated.

Test 1[4]

Directions: Draw a line under the word or phrase that makes the best answer as shown in the sample.

[4] From the Ruch-Cossmann Biology Test, published by the World Book Company, 1924.

Sample: An example of an insect is the
earthworm frog bat cricket sponge lobster starfish

Begin here.

1. Organisms that secure their food supply at the expense of other living organisms are called
saprophytes angiosperms antheridia spirogyra gymnosperms parasites bryophytes 1
2. Dissolved substances enter the roots of plants by means of
capillarity osmosis suction evaporation filtration photosynthesis perspiration 2
3. The male reproductive cell is called the
ovum sepal ovule sperm anther carpel pistil 3
4. A drug that increases the heart action is termed a
narcotic anesthetic bromide poison enzyme stimulant cathartic 4
5. A group of plants characterized by the lack of chlorophyll are the
protozoa algæ gymnosperms fungi bryophytes vertebrates pteridophytes 5

Test 2

Directions: Below are a number of incomplete statements which may be completed by any one of three possible answers. Only one answer is scientifically correct; the other two are partly or entirely incorrect. Study each statement and then make a cross in front of the best answer, as shown in the sample.

Sample: The chief reason why many animals are covered with long, thick fur is that
....The animals are thereby concealed from enemies.
.×.The coat of hair is a protection against cold.
....Such skins are valuable for ladies' furs.

Begin here.

1. The chief function of the white blood corpuscles is the 1
....Destruction of disease germs in the blood.
....Carrying of oxygen to the tissues.
....Carrying of food materials to the cells of the body.

2. Fungous plants are dependent upon other plants for their food supply because 2
....They are so simple in structure.
....Their digestive processes have never been developed.
....The absence of chlorophyll prevents photosynthesis.

3. The act of fertilization is most accurately defined as 3
....The transfer of pollen from stamen to stigma.
....The union of an egg and a sperm cell.
....The maturation of the embryo of an organism.

4. Leguminous plants play a significant rôle in nature because 4
....They will grow on soil too poor to support any other crop.
....Bacteria associated with their roots return nitrogen to the soil.
....The annual value of the hay crop is very great.

Test III is a test of ability to identify parts of structures from drawings. Test IV is a test on the Mendelian laws of inheritance. Test V is a completion test, an example from which is given:

Test 5

Directions: In each of the paragraphs below write in the words that have been left out. Try to find the word for each blank that makes the best sense. The sample has been filled in correctly.

Sample: The *public* ..... *drinking* ..... cup often spreads *disease.* .....

3. Flowers are adapted in three ways to avoid self-fertilization: first, stamens and the ..... may be found in ..... flowers; secondly, the ..... may ripen before the ..... is ready to receive it; and thirdly, the ....., and ..... may be so situated that ..... cannot easily take place.

The reliability of this test is approximately .90. It is valid in construction. The manual of directions is very helpful. The scoring is highly objective and easy except for Test V. Test V, however, is as objective as are most of the completion type tests.

The test is a valuable contribution to standard testing in the science field. It measures principally the achievement objective 1 and, to a more limited extent, objective 2. This test is valuable in much the same manner as the Ruch-Popenoe General Science Test. It can well be used to aid in sectioning classes, to aid in more accurate grading, to pick out pupils who are much in need of help. The test can also be used to measure experimental work with classes. It does not fulfill the requirement of an inventory or diagnostic test, though it may be of some value in each of these functions.

Diagnostic and remedial work in this field has been somewhat limited. While a pupil may be shown to be deficient by the use of these tests and upon investigation may prove to be weak in this subject only, the use of the test does not point out the cause of the weakness. The difficulties of any science for the pupil of average ability may be located within a range of several things. In a science that is new to the pupil vocabulary offers much difficulty. Cotyledon, plumule, ovipositor, and micronucleus are not words in common use. A new science involves a type of "foreign language" to which oftentimes little attention is paid and no drill given. Reading in this new language may be very difficult, and this poor reading ability may form the basis for poor accomplishment in the subject. If laboratory exercises are involved in the work of the class and if these exercises are further complicated with the requirement that drawings be made, many new abilities and techniques are to be developed. Often science work needs much more motivation and application to the life about the pupil than it receives.

CHEMISTRY

Objectives 1, 2, 5, and 6 are most important in the teaching of chemistry. Some good checks or measures of the attainment of high school pupils in the direction of these objectives have been made by the use of the standardized test technique. Powers,[5] in making a study of the subject matter of high school chemistry, made two assumptions. In the first place, he assumed that one of the purposes of instruction is to impart to students a knowledge of the subject matter which is used for instructional purposes, and that, if students fail to acquire such a knowledge, one of the purposes of instruction has failed of accomplishment. His second assumption was that, if students in many different schools are failing to learn the material which is the basis of the instruction, either the subject matter is too difficult or the methods of presentation used in these schools are ineffective. The percentage of students able to do a task correctly after they have had instruction in the subject matter of which the task is representative is a measure of its difficulty. A measure of the difficulty of an item is in a very definite sense a measure of the appropriateness of the subject matter of which the test is representative.

Powers gave objective type tests in fifteen different phases of chemistry, i.e., valence, chemical composition, solubility, completing equations, etc. These tests covered to a marked degree objectives 1 and 2 and in a less degree 5 and 6. The results of this study may be judged by the following:

The items test the students' knowledge of chemical changes and of laboratory and commercial preparations. Here, again, it is observed that the students' knowledge is woefully deficient. Note that nearly 45 per cent do not know the commercial source of ammonia, that nearly 60 per cent do not know the commercial source of glucose, that nearly 75 per cent do not know the

[5] "A Diagnostic Study of the Subject Matter of High School Chemistry," Teachers College, Columbia University, Contributions to Education, No. 149.

ingredients of cement, and that more than 80 per cent do not know the name of the process (hydrogenation) used in the manufacture of Crisco. This finding is particularly disturbing, for one of the largest claims for the functional value of chemistry has been that it gives a knowledge of important industrial processes. These data do not substantiate this claim. It is clear that little such knowledge results from the instruction given in the schools which used this test.

. . . Another claim for the functional value of chemistry is that the student acquired a knowledge of the uses of common chemical substances. That this claim has but little foundation is evidenced by the scores on these items. It seems surprising indeed that nearly 60 per cent do not know that sulfuric acid is used extensively for a dehydrating agent and that two-thirds do not know that cotton is bleached with chloride of lime.

These data reveal to chemistry teachers the futility of much of their labor and suggest the need for remedial measures which will operate to effect economy of time for high school students. Certainly much of their time is wasted if they work with material which is never learned.

Such is the condition in chemistry as revealed by what might well be called a type of diagnostic test.

The material of the objective tests used in this study was later refined and standardized and now makes up the content of the Powers General Chemistry Test.[6] The test is published in two forms, designated as Form A and Form B. Part I of the test is composed of 30 multiple response single choice questions as a test of range of information about chemistry. This part of the test covers a wide range of knowledge in chemistry. An example from Part I is given:

Directions: Part I is a test of range of information about chemistry. In each of the statements there are five choices for the last word or word group. Only one of them will make the statement correct. In each sentence draw a line under the one of these words or word groups which makes the truest sentence.

Sample: Water is a compound of hydrogen and
zinc nitrogen <u>oxygen</u> helium chlorine

[6] Published by the World Book Company, 1922.

The word *oxygen* is underlined because it makes the truest sentence. Following are 30 sentences like this one. Do them in order and complete as many as you can.

Begin here.

1. Substances which hasten a chemical action without themselves undergoing chemical change are called
   catalysts electrolytes ionogens allotrops colloids 1
2. Oxygen was first prepared from chemicals by
   Boyle Priestley Arrhenius Hall Edison 2
3. An essential constituent of all baking powders is
   alum cream of tartar phosphates sodium bicarbonate ammonium sulfate 3
4. Hydrogen fluoride is used for
   bleaching etching glass preserving disinfecting deodorizing 4
5. The acid which is contained in vinegar is
   tartaric lactic acetic oxalic citric 5
6. The sulfur used in commerce
   is mined is made from chemicals is extracted from plants is a product of decay comes out of the ocean 6
7. Acids are characterized by the fact that when in solution they form ions of
   oxygen hydrogen metal non-metal hydroxide 7

Part II of the test consists of 37 items and tests "ability to write formulas and equations, to give chemical names of common substances, and to do simple calculations." A sample from Part II is given:

Directions: Part II is a test of ability to do tasks in chemistry. Directions are given as needed in the body of the test. You are asked to make some calculations. All atomic weights needed are given here. Do not waste any time.

H 1	O 16	N 14	Na 23	Ca 40
S 32	Cl 35.5	Zn 65	C 12	Cu 64

Fill in the right-hand side and balance the following equations:

5. $Zn + H_2SO_4 \rightarrow$ 5
6. $CuO + H_2 \rightarrow$ 6
7. $NaCl + Pb(NO_3)_2 \rightarrow$ 7
8. The formula H_2SO_4 indicates that the elements are in combination in the ratio of .... gram(s) of hydrogen to .... gram(s) of sulfur to .... gram(s) of oxygen 8

Write the molecular formulas for the following:

9. Sodium sulfate 9
10. Hydrogen sulfide..... 10
11. Silver chloride..... 11
12. Potassium carbonate..... 12
13. The equation $Zn + 2HCl \rightarrow ZnCl_2 + H_2$ indicates that .... gram(s) of zinc decompose(s) .... gram(s) of hydrochloric acid and forms .... gram(s) of zinc chloride and .... gram(s) of hydrogen..... 13

The reliability of this test is high enough to give fairly accurate results. The manual of directions is of such a type as to be helpful in the use of the test. The scoring is objective and can be rapidly done. The validity of the test so far as content is concerned is good. The validity as measured by the correlation of the scores of this test with teachers' marks is given in Table 33.

TABLE 33.—CORRELATIONS BETWEEN TEST SCORES AND TEACHERS' MARKS

Cases	*Part I*	*Part II*	*Part I and Part II*
47	.58 ± .05	.64 ± .05	.78 ± .04
135	.51 ± .04	.58 ± .04	.54 ± .04
14	.....	.75 ± .09	.74 ± .09

This table should be interpreted as follows: Part I, correlated with teachers' marks, with 47 cases, gives a coefficient of correlation of .58 with a probable error of .05. Part II correlates .64 with teachers' marks in the same manner. Parts I and II taken together correlate .78 with teachers' marks in Chemistry.

Norms are available by age, grade, and percentile. Since Chemistry is a high school subject, the percentile scores are

of most value, as is explained in Chapter XXII. These are given in Table 34.

TABLE 34.—PERCENTILE NORMS

A student making a score of	64	exceeds 98% of students
A student making a score of	55	exceeds 90% of students
A student making a score of	46	exceeds 75% of students
A student making a score of	36	exceeds 50% of students
A student making a score of	26	exceeds 25% of students
A student making a score of	17	exceeds 10% of students
A student making a score of	8	exceeds 2% of students

Some idea of the value of this test as it may be used in schools is given by Bennett[7] who reports the use of Form *B* of this test in twenty-four high schools with thirty-three teachers of chemistry. The results are given in Table 35.

TABLE 35.—RESULTS BY SCHOOLS

School	*Number of Pupils*	*Median Scores*
1a	38	44
1b	20	38
2	12	22
3a	53	26
3b	21	27
3c	53	37
3d	23	25
3e	31	30
4	19	27
5	9	28
6	56	24
7	87	33
8	50	44
9	55	27
10a	71	36
10b	73	29
10c	51	27
11	117	44
12	9	48
13	51	35
14	54	49
15	16	20
16a	20	37
16b	39	27
17	59	35

[7] "A Study of Pupil Errors in Chemistry," *Journal of Educational Research*, November, 1926, pp. 275-83.

School	Number of Pupils	Median Scores
18a	12	30
18b	30	29
18c	100	23
18d	124	28
19a	30	19
19b	25	29
19c	26	21
20	36	34

The preceding table gives a picture of the difference in attainment by schools as measured by the mean of each class. A detailed analysis of the answers of 1,470 pupils on certain items of Part II was carried out. A picture of the type of errors made was thus obtained. Some examples are given:

1. Write the formula $Mg(PO_4)$ inserting correct subscripts. (The valence of each radical is given.) For $Mg_3(PO_4)_2$, thirty-four per cent wrote 1 for (PO_4); twenty-seven per cent, 1 for Mg; and twenty-six per cent, 2 for Mg.

5. Complete and balance the equation: $Na_2CO_3 + Ca(OH)_2 \rightarrow$. Thirty-one per cent wrote for sodium hydroxide $Na(OH)_2$; thirty per cent wrote $Na_2(OH)_2$; and eleven per cent wrote NaOH instead of 2NaOH.

6. Complete and balance the equation: $Pb(NO_3)_2 + H_2S \rightarrow$. Eighteen per cent wrote for nitric acid HNO_3 instead of $2HNO_3$; seventeen per cent wrote $H_2(NO_3)_2$; and thirteen per cent wrote H_2NO_3.

17. Write the formula for copper nitrate. (Valence not stated.) For $Cu(NO_3)_2$ ninety-seven per cent wrote $CuNO_3$. Other errors were CuN, $CuNO_2$, CuN_2, $Cu(NO_3)_3$, $CuNO_4$, and $CuSO_4$.

27. Name CH_3OH. For methyl or wood alcohol, forty per cent wrote ethyl or grain alcohol; eight per cent wrote methane or methane alcohol; and seven per cent wrote carbon hydroxide. Chromium hydroxide was named by seventeen pupils; glycerine, by fifteen; ammonium hydroxide, by fourteen; carbonic acid, by ten; chloroform, by six; acetic acid, by five; benzene, by three; acetylene, by three; ether, by three; water gas, by two; formaldehyde, by two; and water glass, carborundum, stearine, iodoform dextrine, starch, and vinegar, by one pupil each.

Only 29 per cent of the pupils named this compound correctly.

Further analysis was made of the better pupils (those with scores above the 75 percentile), with this result: "Even among the best students in Chemistry, the valences of the radicals named in the foregoing are none too firmly in mind."

In conclusion Bennett says:

As to the few answers investigated in this study, one is perhaps justified in drawing the following inferences:

1. In formula writing, pupils do not remember accurately the valences of the ordinary simple and compound radicals; they have not become habituated in the processes involved in formula writing; and they have failed, it seems, to learn a simple and rapid check on the correctness of a formula. We as teachers, overwhelmed by the avalanche of subject-matter which a chemistry student is expected to master, have failed to impress the principles involved and to give the drill which might remedy the difficulty in formula writing. If formula writing is essential in the year's work in chemistry, we must *teach* formula writing and be satisfied with nothing less than achievement in that skill. We cannot half teach it without wasting almost *all* the time put into it.

2. From the two cases of equation writing investigated in this study, it seems reasonable to conclude that good instruction and drill in formula writing will go far to help the pupil in writing equations. Furthermore, we have not yet sufficiently appreciated the advantages to be gained by using work equations to introduce formula equations.

3. Concerning a question which requires a chemical name, the formula being given, one's attention is attracted to the data on the answers to No. 27, the name of CH_3OH, because of the wide range covered by the answers. In that long list, sixteen incorrect names for an organic compound are written. Organic chemistry, as now taught, is not functioning so well as it should. Organic chemistry figures in our daily lives to an extent which probably warrants its logical and systematic introduction into the chemistry course. When we teach it as a properly organized unit, we may hope without undue optimism, that our pupils will no longer confuse wood alcohol with glycerine, acetylene, starch, or vinegar.

The study here as in Powers' study, can best be made and sometimes can only be made through the use of the new technique of testing. Such an analysis gives rise to the conclusions just stated and should serve as a challenge to provide better choice of subject matter, better teaching, and more effective learning on the part of the pupils if chemistry is to be taught effectively. The study here has been primarily one of massed data that concerns itself mostly with the groups as a whole. It points out the inadequacy of accomplishment with the present subject matter. A study of the difficulties encountered by particular pupils would reveal much that would point to suitable remedial measures. A study of the psychology of the subject itself, which would more clearly reveal the specific skills and abilities necessary in learning to "complete and balance equation $Na_2CO_3 + Ca(OH)_2 \rightarrow$" as well as other phases of chemistry, would aid decidedly in remedial work.

A test that should be called a type of practice test is the Glenn-Welton New Types of High School Chemistry Tests for Instructional Purposes.[8] This is a series of 36 tests covering each phase of the subject. In all, more than 1,200 items are included in this series of instructional diagnostic tests. Each test is to be used at the completion of the topic studied. These tests will serve in a better manner than any discussed in this field to promote diagnosis and remedial work. The list of tests with time for giving and number of questions is given in Table 36. The tests are objective in type and can be rapidly scored.

TABLE 36.—GLENN-WELTON CHEMISTRY TESTS: TESTS, TIME ALLOWANCE, AND NUMBER OF QUESTIONS

Test	*Time Minutes*	*No. of Questions*
1. Elements, Compounds, and Mixtures.....	15	35
2. Oxygen—Its Properties and Uses.....	30	30

[8] Published by E. R. Glenn and L. E. Welton, 425 West 123rd Street, New York City. These are being revised for a fourth edition.

Test	Time Minutes	No. of Questions
3. Hydrogen—Its Properties and Uses	30	20
4. The Metric System	30	35
5. Measurement of Gases	30	20
6. Transformations of Matter and Energy	30	25
7. Water and Hydrogen Peroxide	30	25
8. Molecular and Atomic Weights: The Atomic Theory	15	35
9. Formulas and Equations	40	25
10. The Chemical Symbols	30	80
11. Carbon, Carbon Monoxide, and Carbon Dioxide	30	35
12. How Atoms Combine: Valence	30	45
13. Nitrogen and the Rare Gases in the Atmosphere	30	27
14. Chlorine and Hydrochloric Acid	40	40
15. Sodium and Sodium Hydroxide	30	45
16. Properties of Solutions: Electrolysis	20	25
17. Compounds of Nitrogen	30	50
18. Types of Reactions	30	30
19. The Compounds of Sulfur	35	50
20. The Periodic Law	20	30
21. The Halogen Family of Elements	40	50
22. Petroleum and the Hydrocarbons	30	20
23. The Fuel Gases	25	30
24. Some Important Coal-tar Compounds	30	30
25. Carbohydrates and Textile Products	30	35
26. Alcohols, Organic Acids, and Foods	30	40
27. Phosphorus, Arsenic, Antimony, and Bismuth	30	45
28. Silicon and Boron	30	30
29. Colloids	20	40
30. The Sodium Family	30	30
31. The Calcium Family	30	35
32. Magnesium, Zinc, and Cadmium	30	30
33. Aluminum and the Silicates	30	30
34. Iron, Cobalt, and Nickel	30	40
35. Copper, Mercury, Silver, Tin, and Lead	30	45
36. Manganese, Chromium, Gold, Platinum, Radium, and Tungsten	20	50

To give some idea of these tests, No. 16 is given:

First fill out the blanks
in the rectangle below:

1. Are you a boy or a girl?.....
2. Age: Yrs..... Mos.....
3. Time of beginning test.....
4. Date.....

Test scored by.....
Test checked by.....
Score = ...

CHEMISTRY TEST NO. 16: PROPERTIES OF SOLUTIONS: ELECTROLYSIS

Read These Directions

Begin with statement No. 1, read it carefully and decide what word, phrase, or number will make the statement complete and correct. Draw a line under the correct word, phrase, or number. Place the number of this answer 1, 2, 3, or 4 in the parenthesis at the right side of the page. Use the margin of the page for any calculations that you wish to make.

Maximum time allowed: 20 minutes.

Write the Numbers Here

Samples:

1. A strong magnet attracts,
   1. Brass; 2. Water; 3. Iron; 4. Wood. ..... (3)
2. The earth revolves around,
   1. The sun; 2. The moon; 3. A comet; 4. A meteor. ..... (1)
3. The earth rotates once in,
   1. 12 hours; 2. 24 hours; 3. 10 hours; 4. 10 minutes. ..... (2)

Write the Numbers Here

1. The solution is said to be ? when it contains more of a substance in solution than is necessary for equilibrium at a given temperature.
   1. Saturated; 2. Super-saturated; 3. Superheated; 4. A soluble. ..... ( )
2. The weight of a substance which will saturate a given volume of a liquid at a definite temperature is defined as a ? of the substance at the stated temperature.
   1. Solute; 2. Super-saturated solution; 3. Solubility; 4. Emulsion. ..... ( )
3. An atom or a group of atoms that carries an elec-

trical charge in a conducting solution is called,
1. An electron; 2. An ion; 3. An electrode; 4. A proton. ..... ( )

4. A typical emulsion is,
1. Lime water; 2. Milk; 3. Hydrogen peroxide; 4. Copper sulfate solution. ..... ( )

5. A solution which conducts an electric current is called,
1. An electrode; 2. An ion; 3. An electrolyte; 4. A colloid. ..... ( )

6. The theory of ionization was proposed by,
1. Faraday; 2. Arrhenius; 3. Davy; 4. Van't Hoff. ..... ( )

7. The liquid which dissolves a substance is called,
1. A solute; 2. A saturated solution; 3. A colloid; 4. A solvent. ..... ( )

8. The electric current passes out of the conducting solution at the,
1. Cation; 2. Anion; 3. Cathode; 4. Anode. ( )

9. The relation of pressure to the solubility of a gas in a liquid was investigated by,
1. Graham; 2. Pictet; 3. Henry; 4. Bridgeman. ( )

10. The laws of electrolysis were discovered by,
1. Smith; 2. Langbein; 3. Faraday; 4. Weston. ( )

11. The terminal at which an electric current enters a conducting solution is called,
1. The cathode; 2. The anode; 3. An ion; 4. The electrolyte. ..... ( )

12. An example of a colloid is,
1. Sugar; 2. Salt; 3. Starch; 4. Copper sulfate. ..... ( )

13. Experiments show that many salts, bases, and acids are,
1. Non-electrolytes; 2. Colloids; 3. Electrolytes; 4. Solvents. ..... ( )

14. According to the theory of ionization an acid is a substance that produces ? when dissolved in water.
1. Chlorine ions; 2. Hydrogen atoms; 3. Hydrogen ions; 4. SO_4 ions. ..... ( )

15. The combination of a negative ion of an acid and the positive ion of a base produces,
    1. A basic salt; 2. A salt; 3. A non-electrolyte; 4. A normal salt. ..... ( )
16. The positive ion in an iron chloride solution will be displaced by,
    1. Tin; 2. Zinc; 3. Copper; 4. Gold. ..... ( )
17. Articles to be plated are always attached to,
    1. The anode; 2. The cathode; 3. The positive electrode; 4. The cation. ..... ( )
18. When a salt solution is electrolized, free chlorine gas and ? can be obtained.
    1. Sodium; 2. Sodium oxide; 3. Sodium hydroxide; 4. Sodium peroxide. ..... ( )
19. The best solvent for a substance like paraffin is,
    1. Water; 2. Alcohol; 3. Gasoline; 4. Vinegar. ( )
20. The cations travel to the ? as the current passes through the conducting solution,
    1. Anion; 2. Anode; 3. Cathode; 4. Positive electrode. ..... ( )
21. In making up a concentrated solution of lime water for use in the home, one should use a good grade of calcium oxide and water at a Fahrenheit temperature of,
    1. 20 degrees; 2. 70 degrees; 3. 80 degrees; 4. 98.6 degrees. ..... ( )
22. Hydrogen can be expelled from water by,
    1. Silver; 2. Copper; 3. Calcium; 4. Arsenic. ( )
23. The positive ion in stannous chloride can be displaced by,
    1. Lead; 2. Gold; 3. Mercury; 4. Zinc. ..... ( )
24. The lead can be displaced in a lead nitrate solution by,
    1. Zinc; 2. Copper; 3. Antimony; 4. Bismuth. ( )
25. The element that is next below hydrogen in the electrochemical series is,
    1. Chromium; 2. Copper; 3. Arsenic; 4. Lead. ( )

Time of finishing test..... Time used = ...minutes.

What Other Students Have Done on This Test

1. The highest quarter of students made a score of ... or higher.
2. One-half of the students made scores between ... and ....
3. The lowest quarter of students made a score of ... or lower.

A.....W.....R.....

Test 16. Score =

PHYSICS

The values that could be attained from the use of reliable tests in Physics, if such tests were available, are stated in detail by Glenn and Heck.[9] The results of testing would point the way for teachers and administrators to carry out the following procedures:

1. Evaluate the aims of instruction in terms of some of the measured results obtained.
2. Decide what items of subject matter are not learned thoroughly enough to be worthy of a place in the course.
3. Make a careful study of the grade placement of physics as a course or to decide which items of subject matter can be taught to the best advantage in general science or in physics.
4. Study the most advisable sequence of topics as represented by the recent high school textbooks.
5. Select drill material to make possible more complete learning of minimum essentials of the subject.
6. Show disinterested educators some of the contributions that physics makes in general secondary education.
7. Determine the usefulness of mental tests in predicting probable school success in high school physics.
8. Test one teaching method against another, such as the project method versus logically arranged subject matter, or the textbook method versus a method which uses a

[9] "Preliminary Studies of Achievement in Physics in Large High Schools." Contributions to Education, Vol. I, Chapter XXX, World Book Company, 1924.

combination of demonstration and individual laboratory experiments.

9. Make studies of different methods of supervising the individual study of pupils, such as supervised study in the classroom versus the study hall, or supervised study in the classroom versus home preparation of lessons.
10. Determine when to release superior students for optional work of an advanced character after certain legitimate standards have been achieved.
11. Determine the extent to which a student profits by classroom instruction with respect to details of the course.
12. Compare the achievement of classes in the same school and in different schools.
13. Make studies concerning the relation of the size of the class to the efficiency of instruction.
14. Be more intelligent in deciding whether a pupil should pass or fail in physics.
15. Compare the achievement of boys and girls.
16. Determine whether prospective college freshmen have profited by the high school course.
17. Classify college freshmen upon the basis of ability in the subject.
18. Show college professors what may reasonably be expected of good high school instruction in physics.

To these it seems could be added (19) Aid in finding and remedying cases of individual difficulties.

The Iowa Physics Test (the tests are, in fact, scales) is an attempt to measure "(*a*) Knowledge of the Fundamental Principles of Physics, (*b*) Ability to Put This Knowledge to Use in Solving Problems One Meets in Ordinary Life." There are three series of tests: Series A, Forms 1 and 2, Mechanics; Series B, Forms 1 and 2, Heat; Series C, Forms 1 and 2, Electricity and Magnetism. Each form is a two-page folder containing 11 or 12 questions. The working time is 45 minutes. The questions are of the recall type: the chance element is eliminated. An example from Series B, Form 1, Heat, of this test is given:[10]

[10] Published by the Public School Publishing Company, 1920.

Value

(9.6) 6. How many degrees Centigrade will 10 calories of heat warm 5 grams of water?

Answer .....

(10.2) 7. The temperature of the air was 45 degrees Fahrenheit at noon and the following morning it was 0 degrees Fahrenheit. What was the drop in Centigrade degrees?

Answer .....

These tests provide an objective measurement of a portion of physics, and can be used for purposes of comparison. Norms (medians and interquartile ranges) are provided in the Manual of Directions. The norms are not sufficiently extensive to throw much light on individual pupil-progress.

The Glenn-Obourn Physics Tests are in form like the Glenn-Welton Chemistry Tests. The twenty-six units for which separate tests have been developed are listed in Table 37.

TABLE 37.—GLENN-OBOURN PHYSICS TESTS: TESTS, TIME ALLOWANCE, AND NUMBER OF QUESTIONS

Test	*Time Minutes*	*No. of Questions*
1. Measurement of length, mass, time and density	15	15
2. Liquid pressure Pascal's law	25	15
3. Air pressure; pneumatic devices	25	20
4. The molecular theory of matter	20	40
5. Force and motion	30	25
6. Work, power, and energy	25	25
7. Mechanics of solids, liquids, and gases (Summary)	30	25
8. Temperature measurements and expansion phenomena	25	25
9. Transformation of work and heat energy	25	20
10. The phenomena of melting and boiling	25	50
11. Conduction, convection, and radiation phenomena	25	20
12. Heat (Summary)	25	25
13. Magnetic phenomena	20	20
14. Static electrical phenomena	25	50
15. Heating, chemical, and magnetic effects of the electric current	20	20
16. The dynamo, motor, and transformer	25	25
17. Magnetism and electricity (Summary)	20	20

Test	Time Minutes	No. of Questions
18. The nature and transmission of sound	20	25
19. The musical scale and musical instruments	25	20
20. Sound (Summary)	20	15
21. The transmission of light	20	15
22. Mirrors, lenses, and optical instruments	25	20
23. The phenomena of color	25	25
24. Invisible radiations; wireless telegraphy and telephone	20	40
25. Light (Summary)	20	30

These tests, covering more than 600 items, offer very much in the way of diagnostic testing as well as accurate methods of marking. They are valuable also in motivating instruction. Much time and effort in physics is given to laboratory work. A measure of the specific outcomes of laboratory work is therefore needed. A composite measure of the expected outcomes of this subject would be exceedingly valuable.

It would be valuable to make a very careful analytical study of the objectives measured by each test in the various sciences. A preliminary outline of this type of study is given in Table 38. It is, of course, to be remembered that, while some objectives are not measured directly or even partially covered in the test, nevertheless they may be measured by a measure of the objectives of teaching of the science with which they are closely correlated. In Table 38, the first six objectives of the teaching of science are listed. Eighteen tests in science are given, and the objectives which each measures are checked. The series starred covers in Science, Chemistry, and Physics, both Aptitude and Training. "X" is used when the objective is measured directly; "p" is used when it is measured partially.

On the whole measurement in the sciences of the junior and senior high school is making some rapid advancement. Perhaps the fact that the particular needs of measurement can be readily seen has much to do with not only the progress that has been made but the progress that may soon be expected.

TABLE 38.—OBJECTIVES MEASURED BY CERTAIN TESTS IN SCIENCE

Name of Test	*1. To give first hand knowledge of the scientific facts, phenomena, principles, and laws of the various sciences as they apply to the lives of people*	*2. To give ability to solve problems using these principles*	*3. To develop scientific attitudes*	*4. To develop scientific interests*	*5. To develop certain technique and knowledge of laboratory procedure*	*6. To do preparatory work for advanced work or direct vocational needs*
Odell Scales for Rating Pupils' Answers in General Science	x	x	x	x		
Powers General Science Test	x					
Ruch-Popenoe General Science Test	x	x			x	
Cooprider Information Exercises in Biology	x					
Jordan-Foran Biology Tests	x				x	
Ruch-Cossman Biology Test	x	x			x	
Gerry Test of High School Chemistry	x	x				
Glenn-Welton New Types of High School Chemistry for Instructional Purposes	x	x			x	
Rich's Chemistry Test	x	x				
Black-Burlingame Test in Elementary Physics	x					
Franzen Test in the Applied Physics of Mechanics, Heat, and Sound	x					
Glenn-Obourn New Types of High School Physics Tests for Investigational and Instructional Purposes	x					
Hughes Physics Scales	x	x				
Peters-Watkins Objective Tests for High School Physics	x	p	p	p		
Michigan Botany Test	x					
Iowa Placement Examinations	x	x				x
Webb Test in Laboratory Resourcefulness					x	
Columbia Research Bureau Physics Test						x

SCIENCE TESTS

GENERAL SCIENCE

Giles-Thomas-Schmidt General Science Examination. By J. T. Giles, S. M. Thomas, and H. W. Schmidt. Published: 1926. Purpose: to measure a pupil's knowledge in the different scientific principles. Range: junior high school. Time to score: short. Standardized: no information.
Publisher: Eau Claire Book and Stationery Co.
Price: $3 per dozen complete sets, postpaid, or 30 cents per single set.

Odell Scales for Rating Pupils' Answers to Nine Types of Thought Questions in General Science. By Charles W. Odell. Published: 1927. Purpose: to measure achievement in general science. Range: junior and senior high school. Time to give: no time limit. Time to score: short. Standardized tentatively.
Publisher: Bureau of Educational Research, University of Illinois.

Powers General Science Test, Form *A*. By S. R. Powers. Published: 1926. Purpose: to test a pupil's knowledge of general science. Range: grades 7-9. Time to give: 30 minutes. Time to score: short. Standardized tentatively.
Publisher: Bureau of Publications, Teachers College, Columbia University.

Ruch-Popenoe General Science Test. By G. M. Ruch and H. F. Popenoe. Published: 1923. Purpose: to measure the accomplishment of pupils in general and elementary science courses. Range: general science classes. Time to give: 40 to 45 minutes. Time to score: 5 to 10 minutes. Standardized.
Publisher: World Book Co.
Price: $1.50 per package of 25 examination booklets, Form *A* or *B*, complete; specimen set, 25 cents.

BIOLOGY

Cooprider's Information Exercises in Biology. By J. L. Cooprider. Published: 1924. Purpose: to test a pupil's knowledge of biology. Range: high school. Time to give: 25 to 40 minutes. Time to score: short. Standardized tentatively.
Publisher: Public School Publishing Co.
Price: 50 cents per package of 25; specimen set, 10 cents.

Jordan-Foran Biology Tests. By E. B. Jordan and T. G. Foran. Published: 1926. Purpose: to test pupil's knowledge of bi-

ology. Range: high school. Time to give: 8 to 15 minutes. Time to score: about 4 minutes. Standardized.
Publisher: The Catholic Education Press.
Price: $1 per package of 25 tests; specimen set, 15 cents.

Ruch-Cossmann Biology Test. By G. M. Ruch and Leo Cossmann. Published: 1924. Purpose: to measure accomplishment in biology. Range: high school and college biology classes. Time to give: 45 minutes. Time to score: 4 to 6 minutes. Standardized.
Publisher: World Book Co.
Price: $1.50 per package of 25 examination booklets, complete; specimen set, 25 cents, postpaid.

BOTANY

Michigan Botany Test. By O. W. Laidlaw and Clifford Woody. Published: 1926. Purpose: to determine the pupil's mastery of botany. Range: first-year botany. Time to give: 25 minutes. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: $1 per package of 25; specimen set, 15 cents.

CHEMISTRY

Glenn-Welton New Types of High-School Chemistry Tests for Instructional Purposes. By Earl R. Glenn and Louis E. Welton. Published: 1922. Purpose: to assist in teaching and to diagnose pupil difficulties. Range: first-year chemistry. Time to give: 15 to 40 minutes. Time to score: 5 to 10 minutes. Standardized.
Publisher: Earl R. Glenn and Louis E. Welton, 425 W. 123rd St., New York City.

Iowa Placement Examinations. By George Dinsmore Stoddard and others. Published: Form *A*, 1925; Form *B*, 1926. Purpose: to measure those factors which contribute to subsequent success in these subjects. Range: college freshmen. Time to give: 40 to 45 minutes. Time to score: 2 minutes. Standardized.
Publisher: Extension Division, University of Iowa.
Price: $3.50 per 100 copies of each series, complete, plus transportation; specimen copy, either form, 5 cents postpaid.

Powers General Chemistry Test. By S. R. Powers. Published: 1922. Purpose: to test achievement in general chemistry. Range: classes in general chemistry. Time to give: Scale *A*,

18 minutes; Scale *B*, 15 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.30 per package of 25 examination booklets complete, either Form 1 or Form 2; specimen set, 20 cents postpaid.

Rauth-Foran Chemistry Test. By J. W. Rauth and T. G. Foran. Published: 1924. Purpose: to measure students' knowledge in chemistry. Range: students of high school chemistry. Time to give: varies according to the exercise. Time to score: about 2 minutes. Standardized.
Publisher: The Catholic Education Press.
Price: $1 per package of 25 tests of either Test I or Test II; specimen set, 15 cents.

Rich's Chemistry Test. By S. G. Rich. Published: 1923. Purpose: to measure attainment in chemistry. Range: high school and college. Time to give: 25 minutes. Time to score: 2 minutes. Standardized.
Publisher: Public School Publishing Co.
Price: $1 per package of 25 tests, either Gamma or Epsilon; Teacher's Manual, 15 cents each; specimen set, 20 cents.

PHYSICS

Camp's Iowa Physics Test. By H. L. Camp. Published: 1920. Purpose: to measure pupils' knowledge of principles in physics and their ability to apply them. Range: first year high school physics. Time to give: 40 to 45 minutes. Time to score: about 1 minute. Standardized.
Publisher: Public School Publishing Co.
Price: Forms 1 and 2, Mechanics, Heat, Electricity and Magnetism, per package of 25 tests, 50 cents each. Teachers' Manual, 15 cents; specimen set, 15 cents.

Columbia Research Bureau Physics Test. By Herman W. Farwell and Ben D. Wood. Published: 1926. Purpose: to measure achievement in physics. Range: high school and college. Time to give: 1 hour, 15 minutes. Time to score: approximately 2 minutes. Standardized.
Publisher: World Book Co.
Price: $1.30 per package of 25 examination booklets; specimen set, 25 cents.

Glenn-Obourn New Types of High School Physics Tests for Investigational and Instructional Purposes. By Earl R. Glenn and E. L. Obourn. Published: 1919, 1920; revised, 1922-23. Purpose: a teaching device as well as a testing device. Range:

high school. Time to give: 10 to 40 minutes. Time to score: about 10 minutes. Standardized.
Publisher: Earl R. Glenn and E. L. Obourn, 425 W. 123rd St., New York City.

The Hughes Physics Scales. By J. M. Hughes. Published: 1926. Purpose: to test the progress of pupils, to diagnose difficulties of pupils, and to rank the pupils in achievement and assign grades. Range: high school physics. Time to give: 20 to 37 minutes, according to the scale given. Standardized.
Publisher: Public School Publishing Co.
Price: 50 cents per package for each division; specimen set, 15 cents.

Iowa Achievement Examinations in College Physics, Series A-2 to A-8 inclusive. By C. J. Lapp. Published: 1926. Purpose: to measure achievement in physics. Range: students of first year college physics. Time to give: 45 minutes for each of series except A-4 and A-8, which are 90 minutes each. Time to score: 2 minutes for each of series except A-4 and A-8, which are 4 minutes each. Standardized.
Publisher: Extension Division, University of Iowa.
Price: $3 per 100 copies of each series, plus transportation; specimen copy of each series, 5 cents postpaid.

Iowa Placement Examinations. By George Dinsmore Stoddard and others. Published: Form *A*, 1925; Form *B*, 1926. Purpose: to measure previous training in certain subjects, and to measure those factors which contribute to subsequent success in these subjects. Range: college freshmen. Time to give: 40 to 45 minutes. Time to score: 2 minutes. Standardized.
Publisher: Extension Division, University of Iowa.
Price: $3.50 per 100 copies of each series complete with directions and scoring keys for either Form *A* or *B*, plus transportation; specimen copy, either form, 5 cents postpaid.

Peters-Watkins Objective Tests for High School Physics. By C. J. Peters and R. K. Watkins. Published: 1926. Purpose: to test knowledge of principles in physics and ability to apply that knowledge. Range: high school. Time to give: 10 minutes for each test. Time to score: 5 minutes for each test. Standardized tentatively.
Publisher: C. J. Peters, University High School, Columbia, Mo.
Price: $20 per 100 booklets; 25 cents per booklet containing all tests.

Tilton's Physics Completion Tests. By Harry Whiting Tilton. Published: 1921 and 1922. Purpose: to test pupils' information in physics and to diagnose their difficulties. Range:

grades 11 to first year of college. Time to give: 20 minutes. Time to score: 5 to 10 minutes. Partially standardized.
Publisher: Harry W. Tilton, South High School, Pittsburgh, Pa.
Price: $1.50 per 100; sample package, 10 tests, 50 cents; sample package, 5 tests, 25 cents; 25 or more copies of one test, 2 cents each.

BIBLIOGRAPHY

Black, N. H. "Some New and Old Types of Physics Tests," *School Science and Mathematics,* Vol. 26: pp. 272-74, March, 1926.

Burgess, T. O. "A Psychological Analysis of Abilities in High School Physics," University of Iowa, Studies in Education, Vol. 3, No. 6, August, 1926.

Caldwell, Otis W. and Committee. *Reorganization of Science in Secondary Schools,* Bulletin No. 26, Department of the Interior, Bureau of Education, Washington, D. C., 1920.

Camp, H. L. "An Evaluation of Standard Tests and Suggested Uses in Improving Physics Teaching," *School Science and Mathematics,* Vol. 23: pp. 441-46, May, 1923.

Camp, H. L. "Scales for Measuring Results of Physics Teaching," *Journal of Educational Research,* Vol. 5: pp. 400-05, May, 1922.

Cooprider, J. L. "Information Exercises in Biology," *School Science and Mathematics,* Vol. 25: No. 8, pp. 807-13, November, 1925.

Cornog, Jacob, and Stoddard, George D. "Predicting Performance in Chemistry," *Journal of Chemical Education,* Vol. 2: pp. 702-08, August, 1925.

Downing, E. R. "Revised Norms of the Range of Information Test in Science," *School Science and Mathematics,* Vol. 26: pp. 142-46, February, 1926.

Downing, E. R. "Norms of Scientific Ability and Achievement in the High School," *School Science and Mathematics,* Vol. 26: pp. 717-21, October, 1926.

Gerry, H. L. "The Need and Use of a Scientific Measure of the Results of the Teaching of Chemistry," *School Science and Mathematics,* Vol. 25: pp. 157-68, February, 1925.

Gerry, H. L. "Types of Tests Desirable for Chemistry and the Present Status of Their Development," *School Science and Mathematics,* Vol. 25: pp. 918-22, December, 1925.

Glenn, Earl R. *Bibliography of Science Teaching in Secondary Schools,* Bulletin No. 13, Department of Interior, Bureau of Education, 1925.

Glenn, H. L. and Heck, A. O. "Preliminary Studies of Achievement in Large City High Schools," *Contributions to Education,* Vol. 1: Chap. 30, World Book Company, 1924.

Herring, J. P. "Derivation of a Scale to Measure Scientific Thinking," *Journal of Educational Psychology,* November, 1919, pp. 417-32.

Hughes, J. M. "The Use of Tests in the Evaluation of Factors Which Condition the Achievement of Pupils in High School Physics," *Journal of Educational Psychology,* Vol. 16: pp. 217-31, April, 1925.

Hurd, A. W. "Observations of Factors Determining Success in Physics," *School Science and Mathematics,* Vol. 25: pp. 121-31, February, 1925; pp. 259-66, March, 1925.

Hyde, W. F. and Guilford, J. P. "Test for Classification of Students in Chemistry," *Journal of Applied Psychology,* Vol. 9: pp. 196-202, June, 1925.

Jones, F. T. "Practice Exercises in Physics," *School Review,* Vol. 26: pp. 341-48, May, 1918.

Malin, Joseph. "Diagnosis by the Use of Science Tests in the Senior High School," *Thirteenth Annual Schoolmen's Week Proceedings,* University of Pennsylvania, 1926, p. 212.

Maxwell, P. A. "Standardizing of First-Year Science Tests," *General Science Quarterly,* Vol. 5: pp. 226-31, 1921.

Powers, S. R. "Achievement in High School Chemistry; and Examination of Subject Matter," *School Science and Mathematics,* Vol. 25: pp. 53-61, January, 1925.

Powers, S. R. "Report of the Committee on Reorganization of the Biological Sciences, Appointed by the Cleveland Biology Teachers Club," *School Science and Mathematics,* Vol. 24: pp. 241-46, March, 1924.

Ruch, G. M. "Tests and Measurements in High School Science," *School Science and Mathematics,* Vol. 23: No. 9, pp. 885-91, December, 1923.

Ruch, G. M. and Cossmann, L. H. "Standardized Content in High School Biology," *Journal of Educational Psychology,* Vol. 15: No. 5, pp. 285-96, May, 1924.

Smith, L. K. "Science Tests: the New Type vs. the Old Type," *School Science and Mathematics,* Vol. 26: pp. 49-52, January, 1926.

Toops, H. A. "General Science Tests," *School Science and Mathematics,* Vol. 25: pp. 817-22, November, 1925.

CHAPTER XIV

Foreign Language

The status of foreign language in our schools may be briefly summarized as follows: (1) Very little if any foreign language is taught in the elementary school (grades 1-6); (2) Foreign language is taught to some pupils in the 7th and 8th grades of the junior high school in the 6-3-3 organization. Usually pupils of more than average ability make up the group that studies foreign language in the junior high school.

The experience of some cities indicates that from forty to fifty per cent of the pupils may profitably take up a foreign language when the opportunity is offered in grades 7 or 8. Others believe that this proportion is too large, and would not advise more than twenty-five per cent of all seventh grade pupils to study a foreign language.[1]

(3) The combined enrollment in all foreign languages is approximately 60 per cent of the total enrollment of pupils in secondary schools (four years). (4) The Latin enrollment in both junior and senior high school is equal to the enrollment of all other foreign languages combined. (5) Some attempts are now being made to develop general try-out and guidance courses in foreign language in the junior high school. (6) A decreasing number of colleges are requiring a certain number of years of foreign language for entrance. (7) For the purpose of developing reading ability in a language, the direct method of teaching is considered by many superior to the translation method.

The objectives of the study of Latin, as given in detail by the Classical League[2] follow:

[1] "Foreign Language in the Junior High School," *Fifth Yearbook*, Department of Superintendence, 1927, p. 294.
[2] *The Classical Investigation*, Part I, pp. 33-35.

1. Ability to read new Latin after the study of the language in school or college has ceased.
2. Increased ability to understand Latin words, phrases, abbreviations, and quotations occurring in English.
3. Increased ability to understand the exact meaning of English words derived directly or indirectly from Latin, and increased accuracy in their use.
4. Increased ability to read English with correct understanding.
5. Increased ability to speak and write correct and effective English through training in adequate translation.
6. Increased ability to spell English words of Latin derivation.
7. Increased knowledge of the principles of English grammar, and a consequently increased ability to speak and write English grammatically correct.
8. Increased ability to learn the technical and semi-technical terms of Latin origin employed in other school studies and in professions and vocations.
9. Increased ability to learn other foreign languages.
10. The development of certain desirable habits and ideals which are subject to spread, such as habits of sustained attention, orderly procedure, overcoming obstacles, perseverance; ideals of achievement, accuracy and thoroughness; and the cultivation of certain general attitudes, such as dissatisfaction with failure or with partial success.
11. The development of the habit of discovering identical elements in different situations and experiences, and of making true generalizations.
12. The development of correct habits of reflective thinking applicable to the mastery of other subjects of study and to the solution of analogous problems in daily life.
13. Increased ability to make formal logical analyses.
14. The development of an historical perspective and of a general cultural background through an increased knowledge of facts relating to the life, history, institutions, mythology, and religion of the Romans; an increased appreciation of the influence of their civilization on the course of western civilization; and a broader understanding of social and political problems of today.
15. Increased ability to understand and appreciate references and allusions to the mythology, traditions, and history of the Greeks and Romans.

16. The development of right attitudes toward social situations.
17. A better acquaintance through the study of their writings with some of the chief personal characteristics of the authors read.
18. Development of an appreciation of the literary qualities of Latin authors read, and development of a capacity for such appreciation in the literatures of other languages.
19. A greater appreciation of the elements of literary technique employed in prose and verse.
20. Improvement in the literary quality of the pupil's written English.
21. An elementary knowledge of the general principles of language structure.

This statement of aims or objectives is quite comprehensive. In the investigation of the Classical League to test the validity of these aims or objectives, scientific studies were made and many tests were developed. To give some idea of the tests that are available and valuable in measuring the objectives listed, the following list is given together with the objective that each test is best fitted to measure:

Objective	
1	Ullman-Kirby Latin Comprehension Test
2	Stevenson-Coxe Latin Derivative Test [3]
3	Carr English Vocabulary Test
4	Thorndike-McCall Reading Test
5	Composition Scale
6	Spelling Test of Latin Derived Words taken from Iowa Spelling Scale
7	Kirby Grammar Test
8	Vocabulary Test Built from Pressey Vocabularies of High School Subjects
9	American Council (Spanish, French or German) Tests
10	None
11	None
12	See E. L. Thorndike, *Journal of Educational Psychology*, Jan. and Feb., 1924, "Mental Discipline in High School Studies"
13	None
14	Davis-Hicks Roman History Tests

[3] W. W. Coxe, "The Influence of Latin on the Spelling of English Words." A doctor's dissertation at the Ohio State University, 1923 (Public School Publishing Company). See Part II, Chap. 1, Sec. 13.

15 Clark-Ullman Classical Reference and Allusion Test
16 Watson's Test of Fairmindedness
17 No test available—could be easily tested
18 Brown Scales for Measuring Growth in Appreciation of Poetry
19 None
20 Schutte English Diction Test; Test of Figures of Speech
21 None

The measurement of achievement in Latin or in any other foreign language may well be compared to the measurement of achievement in reading if objective 1 is to be accepted as sound and desirable as well as the most important. The problems of measurement become problems of the measurement of vocabulary and of oral and silent reading. Much of the measurement of Latin is dependent upon whether the first objective stated is accepted and whether the direct method of teaching is employed. The textbook organization and common practice in teaching indicate, however, that this aim is not generally accepted and that the direct method of teaching is not generally employed. Many Latin tests have been developed and standardized, but most of them are designed for pupils who have received instruction by the indirect method. Such tests tend to assume further, that aims 3, 4, 5, 6, 7, and 8 are the main objectives in Latin teaching. The justification for Latin in the public school system in recent years has been largely on the ground of its value as an aid to better English. It would be natural, then, that objectives 3, 4, 5, 6, and 7 should be the ones stressed. Whether both objective 1 and the group of objectives 3, 4, 5, 6, 7 can be attained by the direct method of teaching in Latin is not as yet definitely known. The content of courses in Latin is now in process of change, due to such factors as junior high school Latin, the inability of beginning second year pupils to translate Cæsar, and new methods of teaching. This change may in a way invalidate the tests now in use. There is only one composite Latin achievement test that may be used successfully to measure in one test several

of the various phases of achievement in Latin. The other tests in this field are on specific phases of Latin achievement.

The New York Latin Achievement Test, which has been recently published, is made up of thirteen parts covering thirteen phases of the study of Latin: vocabulary, English into Latin and Latin into English; vocabulary forms; syllabication and accentuation; verb forms, English into Latin and Latin into English; noun and adjective forms; noun and pronoun meaning and case use; syntactical analysis of English passages; derivation; comprehension of material read in Latin; translation, English into Latin and Latin into English. There is much reason to believe that these tests will fill a need in Latin measurement. One test can be used any time after three months' instruction. Another test is to be used at the end of the year.

The Ullman-Kirby Latin Comprehension Test[4] is a Latin reading test that attempts to measure the ability to comprehend the thought in the Latin passage. The test is somewhat like the Thorndike-McCall Reading Test. The paragraphs, which are of increasing difficulty, are in Latin, and the answers are to be given in English. An answer is credited as right if it shows a correct understanding of that part of the paragraph which is involved in the question. The test has been scientifically developed and now includes:

> Two elementary paragraphs, composed [by the authors] for the purpose, two selections from Cæsar, one from Cicero, one poetical selection, and a third selection from Cæsar, a second from Cicero, a second poetical quotation, and a fourth selection from Cæsar. The stress on Cæsar was due not to the conviction that his works are necessarily the most important, but rather because elementary pupils can, under present conditions, read Cæsar more easily than other classical Latin.[5]

[4] Published by the Extension Division, University of Iowa, 1922.

[5] B. L. Ullman and T. J. Kirby, "A Latin Comprehension Test," *Journal of Educational Research*, Vol. 10: (November, 1924), pp. 311-12.

Two forms of the test are available. The norms, which are given for the end of semesters, and statements of reliability are given in Table 39.

TABLE 39.—ULLMAN-KIRBY LATIN COMPREHENSION TEST

	Semester II		*Semester IV*		*Semester VI*		*Semester VIII*		*Semesters Combined*	
	Form I	*Form II*	*Form I*	*Form II*	*Form I*	*Form II*	*Form I*	*Form II*	*Form I*	*Form II*
Pupils Tested...	123	123	101	101	109	109	90	90	423	423
Median....	11.38	10.15	17.39	18.59	21.81	23.11	25.41	25.67	17.80	19.64
Mean.....	11.20	10.70	17.43	18.19	21.58	23.11	24.85	24.67	18.47	18.76
Sigma.....	2.48	3.96	4.05	3.70	4.79	4.00	4.67	3.99	6.59	6.98
Reliability Coefficient..	.53±.04		.65±.04		.57±.04		.71±.04		.85±.02	

While the reliability of this test is somewhat lower than is desirable, nevertheless it is much higher when more than one semester is included in the group from which the reliability is figured. The test as measured against the regents' examination of New York State gave correlations of approximately .60.

The directions for giving the test and a part of the test are given.

INSTRUCTIONS TO THE PUPIL[6]

To the Teacher: Please read these instructions aloud with the pupils and make sure that all thoroughly understand them.

You will find below a Latin paragraph. Read it through for the thought to see if you can answer the questions which follow. Write the answers in English on the dotted lines. You may refer to the paragraph as often as you like in answering the questions.

[6] From Form 1 of the Ullman-Kirby Latin Comprehension Test, by B. L. Ullman and T. J. Kirby, published by the Extension Division, University of Iowa, 1922.

Practice Paragraph

Duōs fīliōs agricola habēbat, Mārcum et Sextum. Magnus erat Sextus et Mārcus erat parvus. Pater parvum fīlium magis amābat et eum ad urbem saepe mittēbat.

a. Which was more loved by his father, Sextus or Marcus? .....
b. Where did his father send Marcus? .....

Paragraph VI

Read this and then write the answers. Read it again if you need to.

Hīs rēbus gestīs Cūriō ā Bagradā flūmine exercitum Uticam dūcit et prope oppidum castra pōnit. Nōndum opere castrōrum perfectō explōrātōrēs nūntiant hostēs Uticam venīre. Eōdemque ferē tempore prīmum agmen eōrum erat in cōnspectū. Hāc rē Cūriō permōtus praemittit equitēs quī impetum morentur; ipse celeriter ab opere dēductīs legiōnibus aciem īnstruit.

17. Near what town did Curio pitch camp? .....
18. What report did Curio receive? .....
19. For what purpose did he send out the cavalry? .....
20. What were the other soldiers doing before they were drawn up in battle line? .....

The key for scoring the test gives acceptable answers. Other answers are allowed if correct. The scoring is not entirely objective.

Another test of this same type has been developed by R. J. Deferrari and T. G. Foran. This test consists of nine Latin passages from representative authors. The passages increase in difficulty. Pupils answer questions in English on the Latin passages. Equivalent Forms *A* and *B* are available. Complete standards are soon to be published.

There is a marked need for this kind of test. Tests of easy reading and vocabulary type like those found in the elementary grades in silent and oral reading might prove of some special value in beginning Latin. Too little work has been done on this phase of Latin testing.

The White Latin Test is made up of two parts in duplicate forms. Part I is a vocabulary test of 100 well-chosen Latin words. Each word is followed by four English words, one of which is correct. Part II is made up of twenty Latin sentences increasing in syntactical difficulty. Each Latin sentence is followed by four English translations, from which the pupil selects the correct one and marks it. This test measures two important phases of achievement desired in teaching Latin—knowledge of vocabulary and ability to translate Latin into English. The test is easy to give and exceedingly easy to score. The reliability of the test is not fully established. Its validity as to content cannot be questioned. No information is available concerning its validity by comparison with school marks. The norms given in Table 40 show regular and marked progression throughout the eight semesters or four years of high school Latin work. Thirty-five minutes of working time are required for both parts of the test; Part I, 15 minutes, Part II, 20 minutes. The norms for this test follow:

TABLE 40.—TENTATIVE TABLE OF NORMS [7]

Semester	*Vocabulary*	*Sentence*	*Total*
1	29	30	59
2	39	45	84
3	48	57	105
4	56	65	121
5	63	72	135
6	70	77	147
7	76	81	157
8	82	85	167

[7] Published by World Book Company.

The Godsey Diagnostic Latin Composition Test[8] is available in two forms. Each form tests two phases of Latin. These can be best illustrated by an example from the test.

To the Pupil: Draw a circle around the word or phrase which you think is the correct one in each of the Latin sentences in Section I. Also at the right of each Latin sentence draw a circle around the numeral which corresponds to the number of the rule which you think applies to the form you have chosen as the correct one. If in any instance you cannot decide which is the correct form or rule, skip it and go on. Go back to it later if you have time. It is not expected that you will be able to complete all the sentences or indicate all the rules. When you have done all you can with Section I, go on to Sections II and III. You will be allowed 30 minutes in which to work.

Section I

SENTENCES (Draw a circle around the correct form)	RULES (Draw a circle around the number of the rule which applies to the correct form)
a. *The king's sons are leaders of the enemy.* Rēgis (fīlius, fīliōrum, fīliōs, fīlii) ducēs hostium sunt.....	4 7 13 14
b. *They were hurling weapons at the wall.* Tēla in mūrum (coniciēbantur, coniciēbant, coniciunt, conicerent).....	1 8 13 14

Rules

1. Place in which is expressed by the ablative.
2. Place from which is regularly expressed by the ablative with *a, ab, de, e* or *ex*.
3. Accompaniment is regularly expressed by the ablative with *cum*.
4. Possession may be expressed by the genitive.
5. Time when or at which is expressed by the ablative.
6. Necessity or obligation is expressed by the passive periphrastic.
7. The direct object is in the accusative.
8. A clause of purpose has its verb in the subjunctive.

[8] Published by the World Book Company.

9. Means or instrument is expressed by the ablative.
10. The personal agent with the passive voice is in the ablative with *a* or *ab*.
11. Exhortation or command may be expressed by the subjunctive.
12. The gerundive is a passive verbal adjective and must agree with its noun.
13. The subject of a finite verb is in the nominative.
14. Action thought of as going on in the past is denoted by the imperfect.
15. Duration of time is expressed by the accusative.

The tests described give a fairly good picture of the most-used Latin Tests. These tests do not correlate highly with each other. They probably measure the result of the different phases of Latin instruction. All of these tests have some diagnostic value, but no one is highly diagnostic. All of them point out only general tendencies in deficiencies. However, by the making of a diagnostic chart from the tests discussed, individual and class errors can be ascertained. New-type objective examinations over phases of the work just covered can well form the basis for diagnostic work. These are discussed in Chapter XXI.

The Lohr-Latshaw Latin Form Test[9] consists of 35 Latin forms of nouns, pronouns, and verbs. The various elements of the test are of nearly equal difficulty. The pupil taking the test is to indicate the information about each form under the respective headings. An example follows.

This test is given to see how well you recognize Latin "forms." Below is a sample test to familiarize you with the work. Abbreviate your answers by using numerals and by using the first two letters of any words involved. If a verb form does not fit into the four conjugations classify it as irregular, using the abbreviation "Ir" in the column marked "conjugation." There are only three parts of speech involved, nouns, pronouns, and verbs. Write the information about each

[9] Published by the Bureau of Educational Research, University of North Carolina.

word in the space appropriate to its part of speech and on the same line as that on which the word is printed.

SAMPLE TEST

Forms	*Nouns*				*Pronouns*				*Verbs*						*Score*
	Decl.	*Gender*	*Number*	*Case*	*Person*	*Gender*	*Number*	*Case*	*Conj.*	*Voice*	*Mood*	*Tense*	*Person*	*Number*	
vult.....	...	...	...	...	...	...	...	...	Ir	Ac	In	Pr	3	Si	.....
ipsam....	...	...	...	...	3	...	...	...	...	...	...	...	...	...	.....
moneat...	...	...	...	...	...	...	...	...	2	...	...	...	...	...	.....
legionem .	3	...	...	...	...	...	...	...	...	...	...	...	...	...	.....

There are thirty-five forms in the test. The same words are used two or more times with different endings so that this may be a test of your knowledge of *forms,* not of vocabulary.

If you finish before the others, put your pencil down and wait quietly for them to get through. Do not change anything in your paper after you have finished.

When the examiner is sure you understand what to do, he will give the signal to start. Remember, do not look at the other side of this paper until the examiner gives the signal to do so.

The examiner should allow twenty minutes for work on the test.

The test is given on page 347.

A chart showing the types of errors made by the pupils of one class on this test is shown in Table 41, page 348.

The pupils whose initials are at the top of the page should give special attention in study to the items indicated by the letters at the bottom. Pupil A.R. should study especially b. Gender, c. Number, etc.

The chart shown in Table 41 was worked out by the teacher and handed to the pupils as a basis for review. Since a number of pupils needed help in declension, gender, conjugation, and tense, the teacher took these up for review and reteaching in

Forms	Nouns				Pronouns				Verbs						Score
	Decl.	*Gender*	*Number*	*Case*	*Person*	*Gender*	*Number*	*Case*	*Conj.*	*Voice*	*Mood*	*Tense*	*Person*	*Number*	
1. dat.															
2. est.															
3. dabat. . .															
4. sunt. . . .															
5. causam.															
6. hostium.															
7. bellorum															
8. hostem.															
9. videbat.															
10. locus. . .															
11. eos.															
12. fuit. . . .															
13. dixerunt															
14. locum. .															
15. potuit. .															
16. dicit. . . .															
17. quas. . . .															
18. hic.															
19. videt. . .															
20. civitate.															
21. videtur.															
22. pars. . . .															
23. exercitui															
24. dixisset.															
25. vidit. . .															
26. re.															
27. possit. . .															
28. darent. .															
29. putarent															
30. venerat.															
31. veniunt.															
32. dicet. . .															
33. ferat. . .															
34. hac. . . .															
35. posset. .															
														TOTAL SCORE	

class. He then asked those pupils who needed additional help on certain phases to see him at vacant periods during the day. In this way he had an opportunity to do reteaching with those that had special difficulty. Following this, the teacher gave

TABLE 41.—A SUMMARY OF THE TYPES OF MISTAKES MADE BY VARIOUS PUPILS IN THE LOHR-LATSHAW LATIN TEST, SEPTEMBER, 1926

Types of Errors	*Pupils* A.R.	M.D.	S.P.	R.S.	R.B.	A.B.	T.S.	A.H.	M.S.	B.R.	W.R.	L.O.	H.E.	*Pupils Who Should Study the Items Indicated at the Left of the Page*
a. Declension	2	3	1	0	1	2	6	1	2	2	3	0	7	M.D., T.S., W.R., H.E.
b. Gender	5	2	6	1	2	1	1	2	2	6	2	1	4	A.R., S.P., B.R., H.E.
c. Number	5	1	0	0	0	0	0	1	0	0	2	0	2	A.R.
d. Case	3	2	2	1	2	0	0	0	2	0	3	1	3	A.R., W.R., H.E.
e. Person	1	0	1	0	0	0	2	0	1	0	1	0	3	H.E.
f. Voice	1	0	3	0	0	0	0	0	2	1	1	0	0	S.P.
g. Mood	2	0	0	3	2	0	2	0	4	0	2	0	2	R.S., M.S.
h. Tense	3	2	2	6	8	1	2	1	5	6	6	1	4	A.R., R.S., R.B., M.S., B.R., W.R., H.E.
i. Conjugation	3	6	3	8	6	1	7	1	1	4	9	0	1	A.R., M.D., S.P., R.S., R.B., T.S., B.R., W.R.
j. Form	0	0	0	3	2	0	1	0	0	0	0	0	0	R.S.
Total Errors	25	16	18	22	23	5	21	6	19	19	29	3	26	
		a	b	g			a				a		a	
	b			h						b			b	
	c		f	i	h					h				
	d			j	i					i	d		d	
		i	i				i		g				e	
	h								h		h			
	i										i		h	

to each member of the class on successive days a set of practice tests or drill sheets that he had developed, covering a very wide range of Latin difficulties, in conjugation, in tense, in gender, in declension, and in case. The pupils exchanged and scored the papers or practice tests and picked out their own difficulties. These practice sheets were used again later to serve as a review. A retest of the same pupils with the Lohr-Latshaw test revealed an outstanding gain in scores. This was not so significant, however, because much of this gain may have been due to practice effect and also because the test measures only a small part of the expected outcomes of Latin teaching.

The significant phases of this teacher's use of standardized tests are these: (1) he found specific types of pupil difficulty that needed attention; (2) he dealt with those difficulties in an efficient way; (3) he realized the value of measuring before teaching; (4) he realized the inadequacy of the test used as a total inventory of the Latin abilities of his pupils; (5) he found that certain pupils needed little or no review in the various phases of the work tested. While the remainder of the class were taking up their specific difficulties, these pupils read some Latin of an easy type in some works furnished by the teacher. In this manner the objective "to read new Latin" was being furthered with the brighter pupils. (6) The teacher learned something of the technique of making informal objective practice tests. (7) More important than any other outcome was the fact that the pupils worked with a purpose and thereby learned more effectively.

Remedial work in Latin is not possible to a large extent until a careful analysis and diagnosis of difficulties are made. Such a diagnosis of Latin reading as has been carried on by Buswell [10] will give evidence of the specific types of deficiencies of poor readers. Buswell made a study of eye movements in successive readings by various groups of students. In discussing this study in relation to methods of teaching, he says:

[10] *A Laboratory Study of the Reading of Modern Foreign Languages*, The Macmillan Company, 1927.

Students taught by direct method attain in two years' study of a foreign language those fundamental habits which are characteristic of the person who reads. . . . Students taught Latin by the indirect method exhibit those characteristics of reading in no degree at all; rather they exhibit in marked fashion the characteristics of the person who deciphers. When the objective of teaching a foreign language is ability to read, a perfectly clear-cut choice of method is possible. A direct method produces desirable reading habits; a translation method does not. This seems to be perfectly true regardless of the type of language—as applicable to Latin as to French.

The Latin reading test will do much to pick out those with reading difficulty. The cause of this difficulty and methods of overcoming it may best be attacked somewhat as in silent reading in English.

There is a need for an oral reading test in Latin that can serve the same place in diagnosis of Latin reading as the Gray Oral Reading Tests do in diagnosing difficulties in reading. If Latin is to function as a language, speaking, writing, and thinking in terms of it must be practiced. Such suggestions as those given for developing vocabulary in reading might well be included, with some variations, in foreign language study. On this same point the report of the Classical Investigation [11] says:

We recommend that new words should first be met in an enlightening context, oral, written, or printed, and that pupils through daily class-room practice in comprehension at sight should be trained to get at the meaning of most new words through intelligent use of the context with the assistance of such light as is often thrown upon the meaning by related Latin words and by English derivatives.

In studying English derivatives, it has been found quite profitable that specific efforts be made in instruction in Latin to teach these derivatives. Knowledge of Latin words per se does not insure understanding of the English derivatives. This fact applies to Latin words and phrases occurring in English writing and to technical terms found in various subjects.

[11] *The Classical Investigation*, Part One, General Report, 1924.

Much of the syntax of Latin studied in the early part of the work is not widely different from that found in English grammar. A review of English grammatical principles constitutes an excellent introduction to the study of Latin syntax.

The report of the Classical Investigation should be helpful in remedial teaching in Latin.

MODERN LANGUAGE

An excellent group of tests has recently been developed in the Modern Languages most commonly taught in our schools. These tests are developed primarily as a type of survey and accurate measure of achievement in the various Modern Languages—Spanish, French, and German. They have been developed by the American Council on Education for the Modern Foreign Language Study.

The tests in Spanish, German, and French called *Alpha* are alike in form. The German Test *Beta* is not published.

The tests are on the whole highly reliable. They are as valid as any tests are found to be when correlated with teachers' marks. The content is valid and scientifically arrived at. The scoring is objective and can be rapidly done. The tests are available in equivalent Forms *A* and *B*.

Norms for the Alpha Tests and other tests in Modern Language are given for high school students in Table 42.

TABLE 42.—A STATEMENT OF THE RELIABILITY AND THE NORMS OF VARIOUS MODERN LANGUAGE TESTS

Test	*Reliability*	*Semester Norms*							
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
American Council Alpha French Test.....	.89–.93								
Vocabulary		16	22	27	32	37	41	45	49
Grammar		5	12	19	24	29	33	37	40
Silent Reading		6	9	12	15	17	19	20	21
Composition		4.1	5.0	5.8	5.9	6.9	7.8	8.5	7.4
American Council Beta French Test[a].....	.97	41	58	83	123	..	..	..	..

[a] Norms based on study of French as begun in the eighth grade.

Test	Reliability	Semester Norms 1	2	3	4	5	6	7	8
Columbia Research Bureau French Test	.96	..	..	..	150	..	183	..	207 [b]
Twigg French Vocabulary Tests		70	78	83	86	88	90	91	92
American Council Alpha German Test	.82–.89								
Vocabulary		23	25	30	38	42	52	..	..
Grammar		14	17	22	23	24	29	..	..
Silent Reading		8	12	13	17	19	24	..	..
Composition		4	4.3	5	5.2	6	6.4	..	..
Columbia Research Bureau German Test	.96	..	..	..	172	..	203	..	228 [b]
American Council Alpha Spanish Test	.84–.91								
Vocabulary		22	30	37	44	50	55	60	66
Grammar		10	13	16	19	23	26	30	33
Silent Reading		5	9	13	16	19	22	24	26
Composition		3.4	3.7	4.0	4.3	4.6	5.0	5.4	5.9
American Council Beta Spanish Test [c]	.96	43	58	72	103	..	..	..	..
Columbia Research Bureau Spanish Test	.97	..	..	..	164	..	193	..	222 [b]

[b] End of the year norms.
[c] Norms based on study of Spanish as begun in the eighth grade.

Table 42 should be read as follows: American Council Alpha French Test as a whole has a coefficient of reliability of from .89 to .93. The norm of the vocabulary part of the test for the end of one semester is a score of 16, for the end of the second semester 22, etc.

The following discussion and description and examples of the American Council Alpha French Test will illustrate the type of test that has been prepared for the Modern Foreign Language Study, under the auspices of the American Council on Education.

INTRODUCTION [12]

The American Council Alpha French Test was prepared by the Modern Foreign Language Study (1) to measure objectively important aspects of achievement and progress in the ability to read and write French, (2) to make possible more accurate comparisons of attainments in different schools and classes and under different methods and conditions, (3) to serve as a means of classification and placement in securing homogeneity in classes, (4) to be used as instruments of analysis for the investigation of the effects of varying ages, intelligence levels, methods, curricula, and objectives, and (5) to define in more specific terms the immediate objectives of instruction in the study of French.

The test requires two class periods for administration. It is published in two booklets designated as Part I and Part II. Part I of each test measures ability in Vocabulary and Grammar. Part II measures ability in Silent Reading and Written Composition. Two equivalent forms, called Form A and Form B, are available for use.

DESCRIPTION AND DERIVATION OF THE TESTS

The Vocabulary Test consists of seventy-five French words, each of which is followed by five English words. The student shows his knowledge of the word by underlining the one English word which corresponds in meaning to the French word. This five-response technique was adopted because of its objectivity, reliability, and ease of scoring. The words are chosen at regular intervals from each successive fifty words in a word count of 400,000 running words in French.

The reliability coefficient based on a sampling of 600 high school pupils from the first to the fourth year and college students from the first to the fourth semester is +.95. The probable error of a score is 2.03 score points. The reliability coefficients for individual years or semesters range from .84 to .91, with probable errors of measurement of approximately two score points.

The Grammar Test is a test of functional grammar, consist-

[12] From the American Council Alpha French Test prepared by Henmon, Coleman, and Trabue, published by World Book Company, copyright 1927.

ing of fifty items chosen from those common to various texts. Each item has been scrutinized with great care and tested for progressive decrease in difficulty from semester to semester. The testing technique varies throughout the test in accordance with what seems best for each grammatical point.

The reliability coefficients based on various samplings from the four years of high school work range from .89 to .93, with a probable error of measurement of approximately two score points. For four semesters of college work the reliability coefficient is +.87, with a probable error of 2.60 score points.

The Silent Reading Test consists of seven paragraphs of increasing difficulty, with questions on the text in French, the answers to be given in English. The paragraphs were selected from a large number on the basis of their progressive increase in difficulty as shown in preliminary tests. Each question was similarly chosen on a basis of increase from semester to semester in the per cent of pupils passing it.

The reliability coefficient based on a sampling of 174 college students is .85, with a probable error of measurement of 1.41 score points. A later sampling of 464 high school pupils yielded a coefficient of .82, with a probable error of measurement of 1.48 score points.

The Written Composition Test consists of a picture about which the students are required to write a composition. The rating of the composition is done by means of the French Composition Scale (see page 16), prepared for the Modern Foreign Language Study by Professor M. R. Trabue.

DIRECTIONS FOR ADMINISTRATION

The total working time of each part of the test is 40 minutes. At least 5 additional minutes should be allowed for filling the blanks on the first page and reading the directions. Therefore, in preparing to give the test, allow for two periods of 45 minutes each. It is very important that the tests be administered exactly in accordance with the directions. Otherwise the norms given in this Manual do not apply and the scores are therefore of little or no value. For the convenience of the examiner the exact words to be said are printed below and are to be read to the class. Do not attempt to give these instructions from memory. Read the directions to the class slowly and distinctly.

"At the beginning of each test you will find directions explaining exactly how you are to give your answers. Read the directions carefully and notice how the sample questions are answered.

"The time allowed for the Vocabulary Test and for the Grammar Test is given on each test. You may not finish either of these tests within the time limit, but answer as many questions as you can in the time allowed.

"I shall notify you at the end of the given time, and even if you have not finished the Vocabulary Test you are to stop immediately and go on with the Grammar Test.

"If you finish either test before the time is up, look over your work carefully and be sure that every answer is right.

"Now turn the page and begin."

Give no assistance to the pupils while they are taking the test, and see that they get none from each other.

At the end of exactly 15 minutes, say: "All stop work now on the Vocabulary Test and proceed at once with the Grammar Test on page 5. Read the directions first."

At the end of exactly 25 minutes, say: "Stop! Close your booklets."

Vocabulary

Directions: One of the five English words or phrases in each line is a correct translation of the French word. Underscore it, as in Examples A and B.

A. haut hard low high strong brave
B. semaine month week seven seed work

Time allowed: Fifteen minutes.

1. par pair equal by step part
2. trouver pierce trust trouble find join
3. devant near before behind during owing
4. paraitre equal draw prepare leave appear
5. bras arm brave brass air hook

6. souvent memory constant often following soft
7. s'asseoir sit down wake up sleep help hold
8. derrière last behind entrance before stage
9. soin evening thirst silk corn care
10. empêcher prevent destroy cover endure hunt

Grammar

Read carefully the directions before each group of questions. Each answer will be counted right or wrong. When you have completed one group of questions, go on at once to the next.

Time allowed: Twenty-five minutes.

I. Write on the dotted line at the right the correct form of the partitive expression (de, du, etc.):

1. Il demande —— pain (*mas.*). 1. .....
2. Je veux —— viande (*fem.*). 2. .....
3. Il n'y a pas —— pommes (*fem.*) aujourd'hui. 3. .....
4. On voit beaucoup —— enfants (*mas.*) dans les rues. 4. .....

Silent Reading

Directions: Read the first passage attentively twice. Read carefully the four questions that follow. Be sure that you understand each of them. Then show that you have read and understood the whole passage by answering each of the questions *in English*. Your answers should be *brief* but should show that you have read and understood the passage. Treat each passage in the same way.

Time allowed: Thirty-two minutes.

I. Laure a un nouveau livre que son père lui a rapporté de France. Ce livre raconte l'histoire de Jeanne d'Arc. Il est illustré de beaucoup de jolies images en couleurs. Elle lit son nouveau livre et elle regarde les images. Mais le professeur va lui dire d'écrire des devoirs sur son cahier au lieu de lire son beau livre. Elle n'aimera pas cela. Il est plus amusant de regarder de jolies images que d'écrire des devoirs.

1. Comment Laure s'amuse-t-elle?

.....

2. Qu'est-ce que le professeur va lui dire?

.....

3. Pourquoi n'aime-t-elle pas cela?

.....

4. Quel est le sujet de son livre?

.....

TESTS IN FOREIGN LANGUAGE

GENERAL

Iowa Placement Examinations. By George Dinsmore Stoddard and others. Published: Form *A*, 1925; Form *B*, 1926. Purpose: to measure previous training in certain subjects, and to measure those factors which contribute to subsequent success in these subjects. Range: college freshmen. Time to give: 45 minutes. Time to score: 2 minutes. Standardized.
Publisher: Extension Division, University of Iowa.
Price: $3.50 per 100 copies of each series complete, either form *A* or *B*, plus transportation; specimen copy, either form, 5 cents postpaid.

Wilkins' Prognosis Test in Modern Foreign Language. By Lawrence A. Wilkins. Published: 1920. Purpose: to measure the ability of pupils to succeed in the study of foreign language. Range: secondary schools and colleges. Time to give: collective tests, 24 minutes; individual tests, for 25 to 30 students, 45 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.60 per package of 25 booklets, complete; specimen set, 10 cents, postpaid.

LATIN

Carr's English Vocabulary Test. By W. L. Carr. Published: 1921. Purpose: to test pupils' vocabulary. Range: grades 7-12. Time to give: 20 minutes. Time to score: 1 minute. Partially standardized.
Publisher: American Classical League.
Price: 2 cents each.

Deferrari-Foran Tests in Latin Vocabulary and Forms. By R. J. Deferrari and T. G. Foran. Published: 1925. Purpose: to test pupil in Latin vocabulary and forms. Range: high school. Time to give: 35 minutes. Time to score: 5 minutes. Standardized tentatively.
Publisher: The Catholic Education Press.
Price: $1 per package of 25 tests of either form; specimen set, 15 cents.

Deferrari-Foran Latin Tests, Test in Latin Comprehension. By R. J. Deferrari and T. G. Foran. Published: 1925. Purpose: to measure pupils' comprehension in Latin. Range: high school. Time to give: 45 minutes. Time to score: 2½ minutes. Standardized.
Publisher: The Catholic Education Press.
Price: $1 per package of 25 tests; specimen set, 15 cents.

Godsey Diagnostic Latin Composition Test. By Edith Godsey. Published: Form 1, 1922; Form 2, 1923. Range: second to eighth semester of high school Latin. Time to give: 30 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1 per package of 100 tests, complete.

New York Latin Achievement Tests. By H. G. Thompson and J. S. Orleans. Published: 1927. Purpose: to measure achievement in Latin at the end of the first half and at the end of the first year of Latin. Range: first-year Latin students. Test I for the first half year of Latin; Test II for the second half year of Latin. Time to give: 80 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: package of 25 with Manual of Directions, Key, and Class Record, $1.30 net.

Orleans-Solomon Latin Prognosis Test. By J. S. Orleans and Michael Solomon. Published: 1926. Purpose: to foretell what success pupils may be expected to have in the study of Latin. Range: beginning high school Latin. Time to give: 50 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.30 per package of 25; specimen set, 15 cents.

Pressey Test in Latin Syntax. By L. W. Pressey. Published: 1922. Purpose: to diagnose pupil difficulties in Latin. Range: grades 9A-12. Time to give: 15 minutes. Time to score: 2 minutes. Standardized.
Publisher: Public School Publishing Co.
Price: 50 cents per package of 25 tests and 1 direction sheet; specimen set, 10 cents.

Stevenson-Coxe Latin Derivative Test. By P. R. Stevenson and W. W. Coxe. Published: 1923. Purpose: to measure pupils' ability to recognize the Latin in English words. Range: grades 7-10. Time to give: 15 minutes. Time to score: short. Standardized tentatively.
Publisher: Public School Publishing Co.
Price: 50 cents per package of 25 copies, complete, Forms I, II, and III; specimen set, 15 cents.

Stevenson Latin Vocabulary Test. By P. R. Stevenson. Published: 1923. Purpose: to measure knowledge of Latin vocabulary. Range: grades 7-10. Time to give: 15 minutes. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: 50 cents per package of 25 copies, complete, Forms I, II, and III; specimen set, 15 cents.

Tyler-Pressey Test in Latin Verb Forms. By Caroline Tyler and S. L. Pressey. Published: 1922. Purpose: to diagnose pupil difficulties in studying Latin verbs. Range: high school Latin after first semester. Time to give: 15 minutes. Time to score: 2 minutes. Standardized.
Publisher: Public School Publishing Co.
Price: 50 cents per package of 25 tests and 1 direction sheet; specimen set, 10 cents; extra class record sheets, 1 cent each.

Ullman-Kirby Latin Comprehension Test. By B. L. Ullman and T. J. Kirby. Published: 1922. Purpose: to measure pupils' ability to comprehend Latin selections. Range: grades 9-12. Time to give: 30 minutes. Time to score: 2 minutes. Standardized.
Publisher: Extension Division, University of Iowa.
Price: 100 tests containing complete directions, $1.75; specimen set, 10 cents.

White Latin Test. By Dorrance S. White. Published: 1924. Purpose: to measure growth in knowledge of Latin. Range: first four years of Latin. Time to give: 35 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.25 per package of 25 examination booklets, either form, complete; specimen set, 20 cents, postpaid.

FRENCH

American Council Alpha French Test. By V. A. C. Henmon, Algernon Coleman, and Marion R. Trabue. Published: 1927. Purpose: to measure achievement and progress in the ability to read and write French. Range: high school and college. Time to give: 40 minutes for each part. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.25 per package of 25 of each form, complete; specimen set, 35 cents postpaid, for one copy of each form and part, complete.

American Council Beta French Test. By Jacob Greenberg and Ben D. Wood. Published: 1927. Purpose: to measure pupils' achievement in French. Range: junior and senior high school. Time to give: 30 minutes for each part. Time to score: 5 minutes. Standardized.
Publisher: World Book Co.
Price: $1.30 per package of 25, complete; specimen set, 25 cents postpaid, for one copy of each form, complete.

American Council French Grammar Test, Selection Type. By Frederic D. Cheydleur. Published: 1927. Purpose: to measure pupils' achievement in French. Range: high school and college French. Time to give: 22 minutes. Time to score: short. Standardized.
Publisher: World Book Company.
Price: $1.25 per package of 25; specimen set, 20 cents.

Columbia Research Bureau French Test. By A. A. Méras, Suzanne Roth, and Ben D. Wood. Published: 1926. Purpose: to measure pupils' achievement in French. Range: high school and college. Time to give: 90 minutes. Time to score: 5 minutes. Standardized.
Publisher: World Book Co.
Price: $1.30 per package of 25; specimen set, 20 cents.

Handschin's Modern Language Tests: Comprehension and Grammar Test *A:* French. By C. H. Handschin. Published: 1919. Purpose: to diagnose difficulties in French grammar. Range: first year classes in high school French. Time to give: 10 minutes. Time to score: 1 minute. Standardized.
Publisher: World Book Co.
Price: $1 per 50 copies, complete. Specimen set, 20 cents.

Handschin's Modern Language Tests: Silent Reading Test *A:* French. By C. H. Handschin. Published: 1919. Purpose: to measure ability of pupils in elementary French. Range: first two years of French in high school or university. Time to give: 5 minutes. Time to score: 1 minute. Standardized.
Publisher: World Book Co.
Price: $1 per 50 tests, complete; specimen set, 20 cents.

Iowa Placement Examinations. By George Dinsmore Stoddard, G. E. Vander Beke, and C. E. Young. Published: Form *A,* 1925; Form *B,* 1926. Purpose: to measure previous training in certain subjects, and to measure those factors which contribute to subsequent success in these subjects. Range: college freshmen. Time to give: 40 to 45 minutes. Time to score: 2 minutes. Standardized.
Publisher: Extension Division, University of Iowa.
Price: $3.50 per 100 copies of each series complete with directions and scoring keys for either Form *A* or *B,* plus transportation; specimen copy, either form, 5 cents postpaid.

Twigg French Vocabulary Tests. By Alice M. Twigg. Published: 1925. Purpose: to measure the number and value of the French words for which the pupil knows the more common

meanings. Range: first four years of French. Time to give: 30 to 40 minutes. Time to score: short. Standardized.
Publisher: Harvard University Press.
Price: $1.25 per lot of 25, complete (sold only in lots of 25); specimen set, 25 cents.

GERMAN

American Council Alpha German Test. By V. A. C. Henmon, B. Q. Morgan, C. M. Purin, and Stella M. Hinz. Published: 1926. Purpose: to test pupils' knowledge in German. Range: high school and college. Time to give: 45 minutes for each part. Time to score: 5 minutes for each part. Standardized.
Publisher: World Book Co.
Price: $1.25 per package of 25 booklets of each form, complete; specimen set, 35 cents postpaid, 1 copy of each form and part, complete.

American Council on Education German Reading Scales. By M. J. Van Wagenen and S. H. Patterson. Published: 1927. Purpose: to test pupils' ability to associate ideas in German. Range: high school. Time to give: no time limit. Time to score: short. Standardized.
Publisher: Public School Publishing Co.
Price: 75 cents, plus postage, per package of 25 scales and a record sheet.

Columbia Research Bureau German Test. By C. M. Purin and Ben D. Wood. Published: 1926. Purpose: to test pupils' ability to recognize, comprehend, and use German correctly. Range: high school and first three years of college. Time to give: 90 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.30 per package of 25; specimen set, 20 cents.

SPANISH

American Council Alpha Spanish Test. By Milton A. Buchanan, J. P. W. Crawford, and Hayward Keniston. Published: 1927. Purpose: to measure achievement and progress in Spanish. Range: high school and college. Time to give: 45 minutes. Time to score: 4 minutes. Standardized.
Publisher: World Book Co.
Price: $1.25 per package of 25, each part and form; specimen set, 35 cents, complete.

American Council Beta Spanish Test. By Frank Callcott, Robert H. Williams, and Ben D. Wood. Published: 1927. Purpose: to measure pupils' achievement in Spanish. Range: junior and

senior high schools. Time to give: 30 minutes for each part. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.30 per package of 25, complete; specimen set, 25 cents, postpaid.

Columbia Research Bureau Spanish Test. By Frank Callcott and Ben D. Wood. Published: 1926. Purpose: to test a pupil's ability to understand Spanish vocabulary, to comprehend statements, and to translate with grammatical correctness. Range: high school and first three years of college. Time to give: 90 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.30 per package of 25; specimen set, 20 cents.

Contreras-Broom-Kaulfers Test of Silent Reading in Spanish. By M. S. Contreras, E. Broom, and W. Kaulfers. Published: 1927. Purpose: to measure pupils' silent reading ability in Spanish. Range: first three years of high school Spanish. Time to give: 20 minutes. Time to score: short. Standardized: in process.
Publisher: Public School Publishing Co.
Price: not yet determined.

Contreras-Broom-Kaulfers Test of Spanish Vocabulary. By Maria de la Soledad, S. Contreras, E. Broom, and W. Kaulfers. Published: 1927. Purpose: to measure pupils' vocabulary in Spanish. Range: first three years of high school Spanish. Time to give: 25 minutes. Time to score: short. Standardized: in process.
Publisher: Public School Publishing Co.
Price: not yet determined.

Iowa Placement Examinations. By G. D. Stoddard and G. E. Vander Beke. Published: Form *A*, 1925; Form *B*, 1926. Purpose: to measure previous training in certain subjects, and to measure those factors which contribute to subsequent success in these subjects. Range: college freshmen. Time to give: 40 to 60 minutes. Time to score: 2 minutes. Standardized.
Publisher: Extension Division, University of Iowa.
Price: $3.50 per 100 copies, complete with directions and scoring keys for either form *A* or *B*, plus transportation; specimen copy, either form, 5 cents postpaid.

The Wilkins Achievement Tests in Spanish. By Lawrence A. Wilkins. Published: 1926. Purpose: to measure pupil progress in Spanish. Range: first year Spanish. Not standardized.
Publisher: Henry Holt and Co.
Price: $1.20 per package of 25, any test; 10 cents per copy.

BIBLIOGRAPHY

Bond, O. F. "Causes of Failure in Elementary French and Spanish Courses at the College Level," *School Review,* Vol. 32: pp. 276-87, April, 1924.

Briggs, T. H. "Lohr-Latshaw Latin Test," *The Classical Journal,* Vol. 43: pp. 451-65, May, 1923.

Brueckner, L. J. "The Status of Certain Basic Latin Skills," *Journal of Educational Research,* Vol. 9: pp. 390-402, May, 1924.

Byrne, L. "Latin Tests in Iowa High Schools," *University of Iowa Bulletin,* No. 92, 1923.

Cheydleur, Frederic D. "The American Council French Grammar Test," *Bureau of Educational Research Bulletin,* No. 8, University of Wisconsin, 1927.

Coxe, Warren W. *The Influence of Latin on the Spelling of English Words,* Public School Publishing Co., 1925.

Deihl, J. D. "The Basis of Educational Tests in Modern Foreign Languages," *The Modern Language Journal,* Vol. 7: pp. 269-73, March, 1923.

Hanus, P. H. "Measuring Progress in Learning Latin," *School Review,* Vol. 24: pp. 342-51, May, 1916.

Henmon, V. A. C. "A French Word Book Based on 400,000 Running Words," *Bureau of Educational Research Bulletin,* No. 3, University of Wisconsin, 1924.

Henmon, V. A. C. "The Measurement of Ability in Latin," *Journal of Educational Psychology,* Vol. 11: pp. 121-36, March, 1920.

Henmon, V. A. C. "Standardized Vocabulary & Sentence Tests in French," *Journal of Educational Research,* Vol. 3: pp. 81-106, February, 1921.

McPharl, Andrew H. "Freshmen Academic Achievement in College of Students Presenting Four Years of Latin and Those Presenting No Latin," *School and Society,* Vol. 19: pp. 261-62, March 1, 1924.

Méras, Roth, and Wood. "A Placement Test in French," *Contributions to Education,* Vol. 1: Chap. XXV, World Book Company, 1924.

Pettingill, Laura K. "Standardized Tests in Latin," *Classical Weekly,* Vol. XVII: pp. 145-48, March 17, 1924.

Price, Thompson, and Edwards. "English Translation," *School and Society,* Vol. 23: pp. 51-56, January 9, 1926.

Twigg, A. M. "A New French Vocabulary Test," *School and Society,* Vol. 20: p. 734, December 6, 1924.

Ullman, B. L. and Kirby, T. J. "A Latin Comprehension Test," *Journal of Educational Research,* Vol. 10: pp. 308-17, November, 1924.

Wood, Ben D. *New York Experiments with New Type Modern Language Tests,* The Macmillan Company, 1927.

Woodring, Maxie N. "A Study of the Quality of English in Latin Translations," *Contributions to Education,* No. 187, Teachers College, Columbia University, 1925.

CHAPTER XV

Industrial Arts and Vocational Subjects

The tests in the field of Industrial Arts and Vocational Subjects are on the whole quite unsatisfactory, yet some development in measurement is being made in Industrial Arts, Home Economics, and Vocational Trade tests. Mechanical ability or aptitude tests that are somewhat closely connected with this work are discussed in Chapter XX, "Prognosis and Special Ability Tests."

Industrial Arts

Industrial Arts as a special subject is taught ordinarily only in larger school systems. The time allotment for this subject in larger cities is given in Table 43.[1]

Table 43.—Time Allotment in Industrial Arts

	Grades					
	I	*II*	*III*	*IV*	*V*	*VI*
Minutes per week.....	60	60	60	67	70	81
Number of cities.....	18	18	20	22	35	39

When the junior high school is relieved of accrediting relations with the colleges, industrial arts should become a continuous and progressive course of study throughout the three years of the junior high school. Meanwhile industrial arts, in some such general type as household mechanics or general shop, should be made a required subject in the seventh and eighth years and continued as at least an elective in the ninth year. The average time allotment in each of the seventh, eighth, and ninth years approximates three clock-hours a week.[2]

[1] Research Bulletin of National Education Association (1924), Vol. I, No. 5.

[2] Glass, James M. *Curriculum Practices in the Junior High School and Grades 5 and 6,* University of Chicago, 1924.

The teaching of Industrial Arts in the elementary school (grades 1-6) centers around the needs of everyday life in selecting and using the products of the industries with reference to health, economy, art, and social values, leaving training in productive skills to vocational education. Studies of this kind concerning food, shelter, and clothing make up the content of the curriculum. Much effort to correlate this work with geography, history, and language is being made.

The Industrial Arts course of the junior high school is primarily exploratory and prevocational in function. Other exploratory and prevocational courses are Agriculture, Commercial Subjects, Drawing, Home Economics, and Shop. Of these the Shop courses predominate in number in the junior high school. In many places these are short unit courses, i.e., one shop subject such as machine shop or metal work covers only a part of the whole semester's work.

The objectives of the teaching of Industrial Arts are stated as follows. (The first three apply primarily to the elementary school, and the remainder to the junior high school.)

1. To give a basis for judgment in the selection and use of industrial products and service.
2. To develop an appreciation for the economic and social phases of industry.
3. To offer sufficient experience in industrial processes to meet the pupil's need and to illustrate the industry.[3]
* 4. To help, through exploratory courses, in the discovery of special interests and attitudes important for proper educational and vocational guidance.
5. To provide a means for developing technical skill.
6. To provide a means for imparting technical knowledge.
7. To enable the pupil to apply the test of practice to some of his thinking.
8. To develop the mind by providing constructive problems, in materials, which demand a vigorous mental reaction.

[3] A. H. Edgerton, *Industrial Arts in Our Elementary Schools,* Manual Arts Press, 1922.

* 9. To interest in school work, through the concrete application of theory to practice, those pupils to whom the academic studies do not appeal strongly.
10. To create interest in the arts and industries without any reference to their vocational significance.
* 11. To enable the pupil, through participation in coöperative problems, to perform better his duty as a member of his home, his school, and his country.
12. To enable the pupil, through the making of minor repairs and the undertaking of minor construction in the home, to contribute to its economic and material upkeep.
* 13. To serve as an introduction to vocational and prevocational training.[4]

The items marked with an * are very closely related to the statement of purposes or general aims of the junior high school.

Tests in the Industrial Arts field have not been developed to any marked degree. There is a great need for measurement in this field: (a) to determine if the immediate objectives of the teaching of Industrial Arts are attained, (b) to measure the results of different methods of presentation, (c) to define more clearly the aims and the content of these courses, (d) to provide better guidance of pupils, (e) to point out pupil difficulties.

The Partricks Industrial Arts Test[5] was developed to test the pupil's knowledge of diversified activities. A part of the test is reproduced to give some idea of its content.

75. What is the usual voltage of service wires?
2200 1000 50 110
76. What do you use less of to make a lean mixture?
sand cement gravel water
77. What is pica known as?
6 point 30 point 12 point 4 point
78. Which requires the largest floor space?
robin chickadee wren bluebird

[4] "Industrial Arts in the Junior High School," Department of Superintendence, *The Fifth Yearbook*, 1927, p. 406.
[5] Public School Publishing Company.

79. What do you apply to knots before putting on the priming coat of paint?
stain putty shellac varnish
80. For what is spelter a trade name?
metallic iron metallic zinc metallic copper metallic tin
81. What is the evaporation of water from wood called?
veneering wood finishing seasoning warping
82. What sizes of wire would you use for winding the coils of a loose coupler?
8 and 30 20 and 26 30 and 20 20 and 24
83. For what are stocks and dies used?
measuring lumber staining wood cutting threads soldering tin
84. What kind of shelter does a robin prefer?
one with 3 sides with a 1″ entrance with a 2″ entrance with a 3″ entrance
85. Which will warp least?
basswood red gum three-ply wood chestnut
86. What number should be on the shank of a bit used to bore a ¾″ hole?
6 8 10 12
87. To what does "set" refer?
chisel saw knife drill
88. For what is ship lap often used?
studding inside trim rough siding rafters
89. To what does the term "aggregate" refer?
saw milling buying hardware mixing concrete electrical appliances
90. What is the unit of electrical resistance?
watt volt ohm ampere
91. What should you secure, in squaring up stock, before cutting to the proper length?
working face proper width working end working edge
92. What does cement contain?
granite and sand limestone and clay limestone and marble dust sand and ashes
93. To which does field magnet apply?
telegraph key shocking coil wireless motor
94. Which applies to the classification of type?
face lead galley composing rule

95. With what may the end of a 6″ board most easily be squared?
cross-cut saw hack saw plane miterbox and saw
96. What is the usual voltage of line wires?
2200 3000 110 50
97. How are lumber dimensions always given?
thickness-width-length length-width-thickness width-length-thickness thickness-length-width
98. How is electrical power changed into mechanical power?
motor ammeter dynamo transformer
99. Why is the cardboard tube of a receiving set shellacked?
to improve looks to make it stronger to prevent shrinking to improve quality of insulation
100. With which was Marconi connected?
radio telephony printing steamboat wireless telegraph

There is much opportunity for development of Industrial Art work that will test the legitimately expected outcomes of instruction in it. In some special phases of this work new-type examinations are being developed that, if well constructed, will prove of value in Industrial Arts measurements.

HOME ECONOMICS

The growth of Home Economics as a part of the curriculum has been phenomenal during the last twenty years. This subject in the public school system has a rather unique position. The values claimed for it are very much in accord with the larger objectives in education. It may well be a strictly vocational subject. The contributions of this subject are, on the various levels of school work, social, esthetic, utilitarian, prevocational, and vocational. The abilities involved in achievement in this subject vary from specific, definite, and highly specialized skills to broad phases of appreciation.

Measurement in this field has been largely of the specific skills, such as a scale for judging hemming and other phases of sewing, and knowledge of important phases of home eco-

nomics, such as knowledge of foods or household activities. No attempt has been made to test directly such an important achievement as esthetic appreciation of various phases of home activities. It is fair to assume that such appreciation is measured to some extent in tests of information. No really composite test or truly diagnostic or inventory test has been developed. However, good tests are available in the field. A carefully prepared test is the "Home Economics Information Test," [6] by Graduate Students in Home Economics of Teachers College. This test covers a wide range of the Home Economics topics. It is in reality a statement of minimum essentials of attainment of the first eight grades, especially grades 7 and 8. The test comes in three separate sets—Set I, Clothing; Set II, Food; Set III, Household activities. To give some idea of the content of the tests the following table is given (Table 44).

TABLE 44.—HOME ECONOMICS INFORMATION TESTS: TESTS AND NUMBER OF ITEMS

	I. Clothing	
Test 1.	Textiles	21 items
Test 2.	Construction	16 "
Test 3.	Care and Repair	32 "
Test 4.	Selection	35 "
	II. Foods	
Test 1.	Sources of Our Common Foods	20 items
Test 2.	Food Selection	15 "
Test 3.	Food Preservation and Storage	20 "
Test 4.	Laboratory Practices	30 "
Test 5.	Food Values and Health in Meal Selection	25 "
Test 6.	Food Preparation	20 "
	III. Household Activities	
Test 1.	The Girl's Bedroom	12 items
Test 2.	Dining Room	21 "
Test 3.	Dishwashing	11 "
Test 4.	Care of the Kitchen	14 "
Test 5.	Labor-saving Devices	10 "
Test 6.	Home Enjoyment	9 "
Test 7.	Care of Children	12 "
Test 8.	Budget	8 "
	Total number of items in three tests	331

[6] Bureau of Publications, Teachers College, Columbia University.

An excellent analysis of the achievement of 1,000 pupils from 40 widely scattered schools is given by Cooley and Reeves,[7] who have had much to do with the development and use of this test. The analysis made of these results of pupil achievement was primarily of how well the course of study and instruction in Home Economics had attained the legitimately expected outcomes by the end of the eighth grade. As one part of the conclusion this statement is made: "On the whole, the minimum essentials for the completion of the eighth grade, which were set up by the committee, have been justified by the results, the average on the entire test being 79 per cent." This test might well serve several purposes: (*a*) to measure attainment as a basis for promotion, (*b*) to measure the efficiency of a course of study or methods of teaching, (*c*) to provide an inventory test that could be used as an aid in pupil placement or classification for pupils who had previously had some Home Economics work, or as a basis for adjustment in the course to be given, and (*d*) to diagnose pupils' deficiencies.

Parts of an Information Test on Foods are reproduced here.

An Information Test on Foods[8]

The exercises on the following pages are to be done by underlining the statement which will make a true sentence and the most exact statement. Two illustrative exercises are given below. A time limit of 35 minutes for the entire test was made in giving the test for determination of standards.

1. A food rich in fat is
   carrots peanuts beets bread

[7] A. M. Cooley and Grace Reeves, "Some Investigations Concerning the Use of Certain Home Economics Information Tests," *Teachers College Record*, Vol. 24: No. 4, September, 1923.

[8] Prepared by the Test Committee of the Illinois Home Economics Association, A. H. Hess, G. Lord, M. Trilling, and A. B. Robinson. Published by the Public School Publishing Company.

In this exercise, the word "peanuts" is underlined because it will make a true sentence and the most exact statement.

2. Food should be served attractively because it
   stimulates the flow of digestive juices
   saves time
   the teacher advises it
   is more sanitary

In this exercise, the statement "stimulates the flow of digestive juices" is underlined because it will make a true sentence and the most exact statement.

I. Function and Composition

5. A food rich in iron is
   cream of wheat sugar spinach butter
6. The body uses calcium for
   a laxative supplying energy bone building a meat substitute
9. A food rich in vitamins is
   macaroni sugar lean meat leafy vegetables

II. Digestion and Absorption of Foods

12. A foodstuff acted upon by the gastric juice is
    starch cellulose protein mineral matter
13. A foodstuff acted upon by the saliva is
    protein starch sugar cellulose
21. If you brought from home a cheese sandwich and a ham sandwich, a suitable combination for the rest of your school lunch would be
    vegetable soup and baked apple
    baked beans and lemon pie
    doughnuts and rice pudding
    macaroni and chocolate pudding

IV. Food Preparation

23. One should be very careful when adding hot liquid to
    eggs sugar butter vegetables
24. In preparing soft custard, you cook until
    stiff starch is gone it coats the spoon curds have disappeared

25. In making white sauce, you
melt fat, add flour, then add milk
add butter to hot milk, then add dry flour
mix flour with hot milk, then add butter
combine fat, flour, and milk simultaneously

26. The most satisfactory method of making beef broth is to
sear meat and put in boiling water
put meat into cold water and let it come gradually to boiling point
put into boiling water
steam several hours

V. Principles Employed in the Preparation of Different Foods

41. A principle employed in the cooking of mild flavored vegetables is
water lost in evaporation is replaced by soaking in water several hours
mineral matter and flavor are saved by cooking in a small amount of water and serving with vegetables
vegetables retain their shape by boiling gently in a large amount of water
odor and flavor are improved by changing water twice in cooking

42. A principle employed in the cooking of tender cuts of meat is
heat hardens protein so that a crust is formed and juices are retained
mineral matter, extractives and volatile oils are soluble in water
heat and moisture soften and dissolve connective tissue
moderate heat frees fat from the cell walls which inclose it

VI. Leavening Agents

53. Air is used as a leavening agent in
sponge cake baking powder biscuit cookies bread

54. Yeast grows best in bread making when put into
cold liquid lukewarm liquid boiling liquid cold liquid and brought to boil

VII. Food Sanitation

55. A good citizen supports pure food legislation because
it gives him prominence in the community
people can find out the character of the food to be purchased
the people demand it
it is more economical

VIII. Food Economics

59. One would expect to pay less money per pound for
porterhouse steak round chuck neck

IX. Preservation and Storage of Foods

65. The proper conditions for storing cereals are in a
covered glass jar and dry place
covered glass jar and damp place
uncovered glass jar in the ice box
in the pasteboard container in which they came

X. Meal Service and Etiquette

74. After the spoon is used
it should be left in the coffee cup
on the dinner plate
on the saucer
the handle should rest on the tablecloth

XI. Fuels and Cooking Apparatus

76. A fuel saver is
a coal range a gas stove a pressure cooker a thermos bottle

XII. Housewifery

80. In measuring flour, one should
scoop flour from bin with cup
shake down in cup
measure before sifting
measure after sifting

XIII. Food Adjuncts

88. Coffee is made from
chicory seed of a berry tuber nut

XIV. Sources of Our More Common Foods

90. Macaroni is made from
starch wheat barley rye
91. Cereals are made from
roots stems seeds leaves

A domestic science teacher of the ninth grade in a small consolidated school in Indiana began her work with a class of ten pupils feeling that she had no idea of what they had had in previous work. The ninth year course in Home Economics was made up of one-half year in food and one-half year in sewing and textiles. She gave the class of ten pupils the Illinois Test on Foods as a kind of inventory test. She then made a diagnostic chart to determine what phases of the work the pupils were able to do already. The chart is given in Table 45, page 376.

An analysis of the results of this test by the teacher showed two things: (1) that certain phases of the food work were already known to many or all of the pupils; (2) that certain phases of the work offered special difficulty. It was further noted that many of the items in the test that would be covered in a course in foods were also covered by instruction in other subjects or by home training. Take for example, No. 5, "A food rich in iron is cream of wheat, sugar, spinach, butter." The knowledge tested for here may be a legitimate phase of instruction in the Foods course in Home Economics, but it is also commonly taught in health and physiology classes in the public schools. The same is true with question No. 9. Question No. 88, "Coffee is made from chicory, seed of a berry, tuber, nut" might easily be used to test the result of instruction in Geography. No. 54, "Yeast grows best in bread making

TABLE 45.—A PART OF A DIAGNOSTIC CHART SHOWING RESULTS OF INFORMATION TEST ON FOODS GIVEN TO NINTH GRADE CLASS [9]

Name and Age of Pupil	*Question or Item Number*																									*Total Score*
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
J. H. 15 yrs....		x		x			x					x		x	x		x		x	x				x	x	33
E. F. 16 "		x	x				x			x		x	x				x	x	x					x	x	52
L. B. 14 "							x			x		x	x	x					x					x	x	72
T. F. 18 "			x				x							x					x				x	x	x	76
A. F. 14 "	x					x	x					x	x		x				x					x	x	58
E. C. 15 "	x	x	x			x	x			x		x	x	x	x		x		x	x				x	x	53
H. F. 16 "	x					x	x							x	x		x		x					x	x	61
N. S. 14 "				x			x			x		x	x	x	x		x	x	x	x				x	x	57
E. W. 15 "			x	x		x	x			x			x	x	x				x				x		x	57
A. W. 16 "							x					x	x	x	x				x					x	x	69
TOTAL ANSWERS INCORRECT....	3	3	4	3	0	4	10	0	0	5	0	7	7	8	7	0	5	2	10	3	0	0	2	9	10	

[9] There are one hundred items in this test. The chart here gives the diagnostic results only on the first twenty-five items. The total score is also given.

when put into cold liquid, boiling liquid, lukewarm liquid, cold liquid and brought to a boil," could be correctly answered from one's home experience. Another interesting fact revealed by the test was that the girl in the class who was 18 years of age made the highest score. No doubt much of her achievement was due to a background of experience.

The teacher in the case just referred to did two things that are evidences of an intelligent use of test results. She assumed that those items of the test that were missed by one-half or more of the class might well be specially emphasized in the term's work on food. She went through the outline of work for the term and noted those items. She then marked to be omitted from the term's work those items that all pupils knew. She made out review questions covering the items missed by some pupils but by fewer than half the class. The names of the pupils who had missed each item were listed with the review questions. This review, with other organization phases of the work, made up the first two weeks of the term's work. Near the end of the term the teacher gave the test again (a second form was not available). While, as would be expected, a large gain was made by the class as a whole, some few items in the test were still not well answered by the class. These the teacher felt she had not taught well.

There is some danger that the use of a test in this manner might cause the teacher to slight more important phases of the work that were not tested. In this case this was not done. Because of the time saved through the use of this test the teacher had time to allow each pupil to carry some special project in cooking in addition to the regular term's work. This is an example of the manner in which a test used as an inventory measure may be of help to a teacher in her work. No doubt a longer test like the Home Economics Information Test described on page 370 would have better served this purpose.

The Murdock Sewing Scale and the Murdock Analytical Scales represent another type of measure in Home Economics.

They provide a standard scale with which pupils may compare their products and thus measure their own work and note their own difficulty.

The Murdock Sewing Scale is made up of fifteen pictures of samples of sewing. The first sample is marked 0, the eighth sample 7.8, and the fifteenth sample 16.4. Each sample or sheet contains three views of the same sewing sample. One view shows a hem, a seam, hemming, basting, running, backstitch, overcasting, and a combination stitch. A second view shows the reverse side of the seam, and of the overcasting, basting, and backstitch. A third view shows the reverse side of the hem, of the hemming stitch, basting, running, and combination stitch. The numerical values for each entire sample are printed beneath it. Sewing containing these hand stitches, or any of them, can be measured by this scale.

By comparison with the scale the quality of the pupils' work may be measured. The validity of the scale in so far as its content is concerned is good. The correlation of the rating on this scale with teachers' marks is given by Miss Brown [10] for grades 7, 8, 9 as .533.

The scale is of value for two purposes: (1) to place properly pupils transferred from other schools, (2) to pick out early in the term those pupils that need help in hand sewing.

The value of hand sewing in a course in home economics may be questioned; but these measuring scales are valuable if hand sewing is to be taught.

The Murdock Analytic Sewing Scales are made up of five separate picture scales each in running, combination, backstitch, overcasting, and hemming stitches. In one sample that is pictured the stitches are very good; in a second, good; in a third, medium; in a fourth, poor; and in a fifth, very poor. These scales can be used as a self-measuring device by the pupils.

[10] "Investigations Concerning the Murdock Sewing Scale," *Teachers College Record,* Vol. 23 (November, 1922), pp. 459-70.

Trade Tests

Trade tests may be of two kinds: (a) tests of ability to do, of proficiency, or of success in a trade; (b) tests that prognosticate the possibility of proficiency in the trade by measuring capacity or potentiality. In this chapter the former only are discussed. It should be clearly remembered that tests of proficiency attempt to measure in objective, quantitative terms the degree of trade ability already possessed as a result of knowledge and practice in the trade. Tests of proficiency may be tests in which oral or written answers are given. Or they may be picture tests that attempt to develop with pictures as nearly as possible a real situation for testing. Or they may be a performance test in which the shop situation is duplicated as nearly as is practicable.

The army trade tests were somewhat along these lines. For example, if an examination were being given for the purpose of selecting a lathe worker, the person taking the test might be asked to answer orally certain questions about the lathe, its operation, and various methods of using it. In the "Lathe Hand Picture Trade Test" pictures would be shown of the lathe and other tools, instruments, and materials used with it, as well as pictures of its use in various ways. Certain questions would be asked about the pictures that would test the person's knowledge of the machine, its parts, and its use in various ways. An actual performance test with a lathe would consist of a practical tryout on a standardized job. Such a course would require the use of blue prints, materials, and tools characteristic of lathe work. Oral, written, and picture tests measure very largely the amount of theoretical information or knowledge possessed by the one taking the test. The performance test is advantageous in certain kinds of trades where skill is highly important. The performance test has the advantage also of being typical of the trade situation. J. Crosby Chap-

man [11] reports various types of these tests in detail, with instructions for their use.

Standard trade tests have not been well developed or widely used except during the World War. It is probable that much can be done in this field by a combination of the efforts of the tradesmen and those that specialize in the technique of test building. Standard trade tests have found very little use in vocational schools. This has been due in part to two things: first, because of the objectivity of the achievement in many phases of vocational work, the need for standardized measures is not great; second, vocational education as viewed by public school administrators has not aimed to develop skilled tradesmen. Vocational educators are probably more committed to the development of tradesmen in vocational education than are general school administrators. It is probable that there is much need, however, for standard trade tests as a measure of achievement of the vocational school work. Trade unions have developed tests for use in many phases of trade work. These tests are used in some vocational and trade schools, and many of them are to be highly commended.

TESTS IN INDUSTRIAL ARTS AND VOCATIONAL SUBJECTS

HOME ECONOMICS

Illinois Information Test on Foods. By the Test Committee of the Illinois Home Economics Association. Published: 1924. Purpose: to measure knowledge of foods. Range: high-school students in home economics. Time to give: 35 minutes. Time to score: 10 minutes. Standardized.
Publisher: Public School Publishing Co.
Price: 75 cents per package of 25; specimen set, 10 cents.

Murdoch Analytic Sewing Scale for Separate Stitches. By Katherine Murdoch. Published: 1922. Purpose: primarily for pupils to use in measuring their own work. Standardized:

[11] *Trade Test*, Henry Holt and Company, 1921.

norms for the Murdoch Sewing Scale may also be applied to this scale.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: 25 cents.

Stiebeling-Worcester Chart for Diagnosing Defects in Buttonholes. By Hazel K. Stiebeling and Dean A. Worcester. Published: 1922. Purpose: to diagnose defects in buttonholes. Time to give: no time limit. Standardized.
Publisher: Bureau of Educational Measurements and Standards, Kansas State Normal School.
Price: 17 cents postpaid; 3 or more copies, 15 cents each.

Streeter-Trilling Test at the Completion of Food Preparation Courses in the Junior High School. By Nina Streeter and Mabel B. Trilling. Published: 1927. Purpose: to test pupils' ability to judge the best way of preparing foods. Range: junior high school. Time to give: 17 minutes. Time to score: short. Standardized: no information.
Publisher: Public School Publishing Co.
Price: 75 cents, plus postage, per package of 25 tests and direction sheet.

INDUSTRIAL ARTS: WOODWORKING

Nash-Van Duzee Industrial Arts Test: Woodwork, Scale *A*. By Harry B. Nash and Roy R. Van Duzee. Published: 1927. Purpose: to measure technical and related information covering the subject of woodwork. Range: junior and senior high school. Time to give: 40 minutes. Time to score: short. Standardized.
Publisher: Bruce Publishing Co.
Price: $1.25 per 25 tests; specimen set, complete, 25 cents.

Wells Knowledge Test in Woodworking. By George K. Wells. Not yet published. Purpose: to test the knowledge side of elementary woodwork. Range: high school. Time to give: 30 minutes. Time to score: 10 minutes. Standardized.
Publisher: George K. Wells. Box 242, Terre Haute, Ind.
Price: 25 cents per copy.

VOCATIONAL, PRE-VOCATIONAL, AND VOCATIONAL GUIDANCE

Hunter Shop Tests. By William L. Hunter. Published: 1927. Purpose: to test pupils' knowledge concerning processes, tools,

and materials. Range: students studying manual arts. Time to give: 5 to 8 minutes. Time to score: short.
Standardized: in process.
Publisher: Manual Arts Press.
Price: 25 cents for 25 copies of any one of the tests; specimen set, one of each test, 25 cents.

MacQuarrie Test for Mechanical Ability. By T. W. MacQuarrie. Published: 1925. Purpose: to test general mechanical ability. Range: age 10 or older. Time to give: 35 minutes. Time to score: 8 minutes. Standardized.
Publisher: Research Service Co.
Price: Form I, $1.50 per package of 25 or $6 per 100 copies; specimen set, 50 cents for 5 copies with directions.

Wells Knowledge Test in Machine Shop. By George K. Wells. Purpose: to test the knowledge side of machine shop work. Range: high school. Time to give: 30 minutes. Time to score: 10 minutes. Standardized.
Publisher: George K. Wells, Box 242, Terre Haute, Ind.
Price: 25 cents per copy.

Wells Knowledge Test in Printing. By George K. Wells. Purpose: to test the knowledge side of elementary printing. Range: high school. Time to give: 30 minutes. Time to score: 10 minutes. Standardized.
Publisher: George K. Wells, Box 242, Terre Haute, Ind.
Price: 25 cents per copy.

MECHANICAL DRAWING

Wells-Laubach Knowledge Test of Mechanical Drawing. By George K. Wells and M. L. Laubach. Purpose: to test the knowledge side of high school mechanical drawing. Range: high school. Time to give: 30 minutes. Time to score: 10 minutes. Standardized.
Publisher: George K. Wells, Box 242, Terre Haute, Ind.
Price: 25 cents per copy.

BIBLIOGRAPHY

Anderson, L. D. "Objective Measurement in Shop Courses," *Industrial Arts Magazine*, Vol. 15: pp. 263-67, August, 1926.

Bauersfeld, A. G. "Application of Standard Tests and Measurements in the Manual Arts," *Industrial Arts Magazine*, Vol. 13: pp. 14-19, January, 1924.

Bowman, Leona F. and Trilling, Mabel B. "Bowman and Trilling Information and Reasoning Test in Textiles and Clothing," *Home Economics in American Schools,* School of Education, University of Chicago Monograph.

Brown, C. M. "Construction and Use of Information Tests in Home Economics," *Journal of Home Economics,* Vol. 16: pp. 251-56, May, 1924.

Brown, C. M. "Investigations Concerning the Murdoch Sewing Scale," *Teachers College Record,* Vol. 23: pp. 459-70, November, 1922.

Brown, C. M. "What Can Educational Measurements Do for Home Economics," *Journal of Home Economics,* Vol. 16: pp. 191-96, April, 1924.

Conley, E. "Use of Measurements and Objective Tests in Home Economics," *Proceedings of National Education Association,* 1925, pp. 902-05.

Cooley, A. M. and Reeves, G. "Some Investigations Concerning the Use of Certain Home Economics Information Tests," *Teachers College Record,* Vol. 24: pp. 374-92, September, 1923.

Duwelius, A. W. "Measuring Mechanical Drawing Scale," *Industrial Education Magazine,* Vol. 29: pp. 16-17, July, 1927.

Gleason, C. W. "An Experiment in Testing Engine Lathe Aptitude," *Journal of Applied Psychology,* Vol. 10: pp. 171-79, June, 1926.

Hunter, W. L. "Tests for Shop Courses," *Industrial Arts Magazine,* Vol. 15: pp. 29-30, January, 1926.

Jones, Edward S. "The Wooley-Test Series Applied to the Detection of Ability in Telegraphy," *Journal of Educational Psychology,* Vol. 8: pp. 27-34, January, 1917.

McDonald, D. J. "Tests in Manual Subjects," *Industrial Arts Magazine,* Vol. 11: pp. 125-30, April, 1922.

Murdoch, K. "The Measurement of Certain Elements of Hand Sewing," *Contributions to Education No. 103,* Teachers College, Columbia University.

Murdoch, K. "A New Analytic Sewing Scale," *Teachers College Record,* Vol. 23: pp. 453-58, November, 1922.

Myers, C. E. "Relation of Measurement to Vocational Education in Agriculture," *Proceedings of the National Education Association,* 1925, pp. 698-901.

Nash, Harry B. and Van Duzee, Roy R. "Standard Tests in Industrial Arts; Woodwork," *Industrial Arts Magazine,* Vol. 16: pp. 248-50, July, 1927.

Ray, R. S. "Constructing a Scale for Grading Mechanical Drawing in the High School," *Industrial Arts Magazine,* Vol. 9: pp. 431-33, November, 1920.

Smith, Z. M. "A Test of Animal Husbandry," *Thirteenth Annual Conference on Educational Measurements,* Indiana University, School of Education, 1926.

Toops, H. A. "Effect of Tests and Measurements in Shop Work on Shop Instruction," *Industrial Arts Magazine,* Vol. 15: pp. 400-01, November, 1926.

Toops, H. A. "Theoretical Principles Underlying Tests and Measurements in Vocational Education," *Proceedings of the National Education Association,* 1925.

Trilling, M. B. and Williams, F. "Standardized Tests in Textiles and Clothing," *Journal of Home Economics,* Vol. 12: pp. 486-91, November, 1920.

Ward, C. M. and Toops, H. A. "Performance Test of Ability in Using Measuring Tools," *Industrial Education Magazine,* Vol. 27: pp. 177-80, December, 1925.

Webster, L. J. "Rating Mechanical Drawing," *Industrial Arts Magazine,* Vol. 15: pp. 176-77, May, 1926.

Williams, Florence and Knapp, Clara B. "Williams and Knapp Scale for Measuring Skill in Machine Sewing," *Home Economics in American Schools,* Department of Education, University of Chicago Monograph.

Woellner, R. "Locating Difficulties in the Learning Process in Practical Arts Education," *Industrial Arts Magazine,* Vol. 15: pp. 341-43, October, 1926.

CHAPTER XVI

Commercial Subjects

Commercial subjects have come into the high school during the last twenty-five years. Their growth as measured by enrollment has been phenomenal. With the development of the junior high school as a unit of organization, courses in commercial subjects have been made an integral part of the curriculum.

The junior high school curriculum tends to include the commercial course primarily for two functions, namely (*a*) to develop try-out courses through which to perform the guidance function, and (*b*) to prepare for the immediate needs of junior commercial workers. This second function has as its aim the giving of specific preparation for work to those pupils who leave school before they have finished a high school commercial course, and who enter the business world therefore as junior workers in such occupations as clerks, typists, etc. The senior high school, on the other hand, attempts to train classes of workers such as stenographers, bookkeepers, clerks for general office work, secretarial workers, and salesmen.

The commercial work of the junior high school, which is largely given in grades 8 and 9, consists of such courses as bookkeeping, business arithmetic, business practice or office activities, typing, and shorthand, in addition to the regular spelling, penmanship, and English. At present shorthand is being taken out of the junior high school and given in the commercial work only of the senior high school. There is an attempt to fuse as many of these courses as possible into one. The commercial work of the senior high school is made up of such

special subject courses as stenography, typing, commercial arithmetic, bookkeeping, commercial geography, office practice, besides the English courses that are specialized to conform to commercial needs.

Measurement in the field of commercial work may be of two types: (1) Measurement of achievement or proficiency in the subject, and (2) measurement of the prognosis type that attempts to measure potentiality. Only the former will be considered in this chapter, as the prognosis type is discussed in the chapter dealing with prognostic tests. In some ways these tests of proficiency and of potentiality overlap in content. Measurement of achievement in certain commercial subjects is very closely associated with measurement of achievement in elementary school subjects. A good stenographer must know, in addition to the shorthand taught in the commercial course, how to spell the words she takes in dictation, how to punctuate sentences, how to paragraph, and in some cases how to write well. Many of these minimum requirements of stenography are stated as objectives in the course of study of the elementary school. If these minimum essentials are attained in the elementary school, much less is to be accomplished in the high school commercial course.

What has been said here of stenography is true of typing. In bookkeeping the requirements are (1) accuracy and speed in handling the four fundamental processes and short-cut methods with fractions and decimals, (2) ability to keep records on business blanks, some of which are complicated, (3) ability to analyze and classify accounts as in journalizing or in the posting of the ledger. The elementary school can well contribute very much to the first of these minimum requirements. Any measurement of achievement in bookkeeping is at least in part a measurement of achievement in elementary school arithmetic. Because of this close association of previous training in elementary school with the requirements of the commercial subjects, inventory tests are highly desirable and should be developed

and widely used in the training of commercial workers. Not much progress, however, has been made in this line.

The tests in commercial subjects here considered are tests in typewriting, shorthand, bookkeeping, and commercial arithmetic. Tests in salesmanship and secretarial work are not as yet available.

The National Business Ability Tests, developed under the direction of Sherwin Cody [1] and with the advice of Dr. George D. Strayer, consist of: (1) Test of Tabulation (Mental Alertness); (2) Reproducing Instructions; (3) Test on Invoicing; (4) Test on Fundamentals of Arithmetic; (5) Test in Business Arithmetic; (6) Test in English which consists of tests of spelling, punctuation, grammar, and business usage required for good letter writing; (7) Elementary Test in Letter Writing; (8) Test on Answering Letters (advanced); (9) Stenographic Tests; (10) Tests on Copying for the Mimeograph; (11) Test on Addressing Envelopes with a Pen; (12) Test of Filing. These tests comprise a basic set of tests in the field that may well serve as a suggestion for better work of this kind.

TYPING

A good test of proficiency in typing is the test developed by E. G. Blackstone. The test is available in five equivalent forms, *A, B, C, D,* and *E*.[2] A great deal of care has been taken to make the material of these various forms highly comparable. The test attempts to measure proficiency in typing by a careful measure of speed and accuracy in the copying of a typewritten letter. The pupil is given three minutes in which to write as much of the letter as possible. The final score is based on the number of strokes per minute calculated by means of the copy, and on the number of errors made.

[1] Sherwin Cody, *Commercial Tests and How to Use Them,* World Book Company, 1919.

[2] Published by the World Book Company, 1923.

Dear Sir:
2 4
Answering your request, we
7 9 12 13
recently quoted you a low price on our
16 18 19 20 21 23 24 26
pianos, and not having received your
28 30 31 33 36 38
order, we take occasion to inquire
40 41 43 46 47 50
whether you have as yet bought. If you
52 54 55 56 58 60 62 63
have, will you tell us what make you
65 67 68 70 71 72 74 75
selected and in what respect our offer
79 80 81 83 85 87 89
failed to interest you?
91 92 95 97
We make this request only as a
98 100 102 105 106 107 108
matter of business, for we are making
110 111 114 116 117 118 121
our plan just as attractive to the
122 124 125 126 130 131 132
customer as we know how, and the
135 136 137 139 141 142 143
rapidly increasing number
146 150 152

The number on this copy underneath the last word written by the pupils is the number of strokes per minute. Disregard any unfinished word. If letters are repeated or omitted, corrections are made accordingly (directions are given in the manual). To determine the number of errors, draw a circle around each error such as an incorrect letter, punctuation, spelling, spacing, and other errors. To get the score use the following formula:

$$\text{Score} = \frac{\text{Strokes per minute} \times 10}{\text{Errors} + 10}$$

Papers can be scored by the pupils. The following example is from a pupil's paper.

Dear Sir:

Answerinf your requiest, we recently quoted you a low price on our pianos, and nor having received your order, we take occasion to inquire whether you have as yet bought. If you have will you tell us what makeyou selected and in what respect our offer failed to interest you?

We make this reqyest only as a matter of business, for we are making our plan just as attractive to the customer as we know how, and the rapidly increaseing number.....

The last word written in three minutes is "number" in the second paragraph of the copy. This gives the strokes per minute as 152. The number of errors is six. Substituting in the formula score $\frac{152 \times 10}{6 + 10} = \frac{1520}{16} = 95$. This pupil has had six months of instruction in typing. Comparison of this score with the norms given in Table 46 shows that her work is slightly better than the five-months norm.

TABLE 46.—BLACKSTONE TYPEWRITING TEST: NORMS

Months of Instruction	*Norm*
5	88
10	148
15	178
20	200
25	220
30	236

The value of the test is twofold. The results of a measure of this kind, that can be given as often as six times in a term, can be charted in such a way as to motivate the work and bring the pupil to the point of grading his own work. Because of the care with which the material of this test has been selected, a study of the pupils' errors may serve to point out difficulties to be overcome. The manual, which includes the test, gives many helps in the use of the results of the test. If an additional number of equivalent forms of this test were developed so that one form of the test might be used each month during the first year of typing instruction, the test would serve an even better purpose than it now does as an aid in classroom instruction.

Another test in typewriting that is of much value is the Thurstone Examination in Typing. Three phases of typing are measured by it: (*a*) ability to type corrected typewritten copy, (*b*) ability to tabulate and type items written in long hand, (*c*) ability to find and correct misspelled words. These phases are more closely related to the usual work of a stenographer than are the phases measured by the Blackstone Test. Scores made on this test are good measures of the general typing ability of the pupil. The value of the test is lessened because of the lack of duplicate forms.

SHORTHAND

A very important part of the equipment of a stenographer is the knowledge of, and ability to use, shorthand.

Measurement in shorthand is in some ways as complete as is measurement in any other subject of the secondary field. The abilities in shorthand are (*a*) ability to take dictation, which consists of ability to translate oral words into shorthand and ability to write this shorthand with speed and legibility; (*b*) ability to translate shorthand notes into English.

E. R. Hoke has performed an excellent service in this field

of testing with the development of measuring instruments for various phases of shorthand ability such as knowledge of shorthand characters, speed of writing shorthand, and ability to read shorthand. His work is based on the Gregg System of shorthand. Some of these tests are illustrated or the various abilities tested are briefly described on pages 391-395. To the 1,000 words of the Ayres list of most common words were added 500 most frequent phrases or combinations of words for which the Gregg System has symbols. This constitutes a list of 1,500 most common words and phrases for which shorthand is used. These words and phrases make up the Measuring Scale for Knowledge of Gregg Shorthand. This scale is in form very much like the Ayres Spelling Scale. These words and phrases also form the content of 100 duplicate vocabulary tests.

Shorthand Penmanship

A measuring scale for Gregg Shorthand Penmanship is much like a handwriting scale of legibility. The following directions are given:

DIRECTIONS [3]

Write the following sentence on the blackboard *in shorthand:*

1 2 3 4 5 6 7 8 9
"Fourscore and seven years ago our fathers brought forth
10 11 12 13 14 15 16 17 18 19
upon this continent a new nation, conceived in liberty, and
20 21 22 23 24 25 26 27 28 29
dedicated to the proposition that all men are created equal."

Have the pupils write this three times for practice, to familiarize them with the text and the shorthand. Collect these papers.

See to it that all have paper and two pencils (or fountain pen). Then say: "Do not write anything until I give you the signal to start. You will have three minutes in which to write

[3] Published by the Gregg Publishing Company, 1922.

the sentence as often as you can. Be careful to write as well as you can. You will be rated both for speed and for quality.

TABLE 47.—CLASS RECORD SHEET[4]

Score for Quality	Score for speed in words per minute																	Totals
	0- 10	11- 20	21- 30	31- 40	41- 50	51- 60	61- 70	71- 80	81- 90	91-100	101-110	111-120	121-130	131-140	141-150	over 150	
95																		
90																		
85																		
80																		
75																		
70																		
65																		
60																		
55																		
50																		
45																		
40																		
35																		
30																		
25																		
under 25																		
TOTALS																		

Start and stop promptly when the signals are given. All ready. Start!" Keep the time very carefully, using a watch with a

[4] Copyright, 1922, by the Gregg Publishing Company.

second-hand, allowing *exactly* three minutes. "Stop!" Collect the papers.

The score for speed is ascertained by multiplying 29 by the number of times the sentence has been completely written, adding the number of the last word written. This is the total number of words written and should be divided by three to get the score for speed in words per minute. Write this score in the upper right-hand corner of the paper.

When all have been scored in this way, score the papers for quality by the use of the scale, writing this score under the score for speed.

Enter the scores on the class record sheet by making a tally-mark in the square under the proper score for speed and opposite the score for quality. Total these scores both ways.

The class record sheet is given in Table 47. By drawing on this record sheet heavy lines representing standards in speed and accuracy and then putting in the pupils' scores by tallying, it is possible to determine the standing in relation to both qualities. (See pages 188 and 222.)

Stenography Reading

Ability to read shorthand is measured by having the pupil read shorthand. Comprehension of shorthand is determined by a rather unique method. In the shorthand material to be read every tenth word is in English. Two English words are given the pupil, who must choose from the two words the one which is correct. If the pupil chooses the right word, it is assumed that he comprehends the shorthand that comes before it. Error due to chance guessing is corrected in scoring. The score is the number of shorthand words read correctly per minute.

The Diagnostic Shorthand Tests [5] are made up somewhat in the same manner as the Hoke tests. They attempt to measure (*a*) Knowledge of principles, (*b*) Penmanship ability, (*c*)

[5] Ethel A. Rollinson, "Manual Diagnostic Shorthand Tests," Gregg Publishing Company, 1924.

Facility of writing, (*d*) Reading ability. In the Manual for this test are given many suggestions for diagnosing difficulties and overcoming them. These tests are in some respects an improvement over the Hoke tests. There are four tests of this group, each of which measures the work covered by one of the first four chapters of the Gregg Shorthand Manual—Test I for Chapter I, Test II for Chapter II, and so on. An example of the test of reading ability follows.

Test I, Reading Ability Test of the Diagnostic Shorthand Tests by Ethel Rollinson.

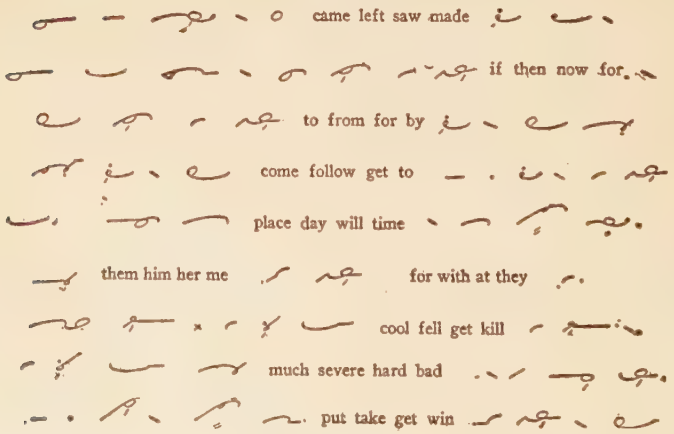
Reading Ability:—On the fourth page is a test written in shorthand with an occasional group of four longhand words. Only one word will complete the sense. Read the shorthand, determine the longhand word which will complete the sense, and underscore the word. You will work for five minutes.

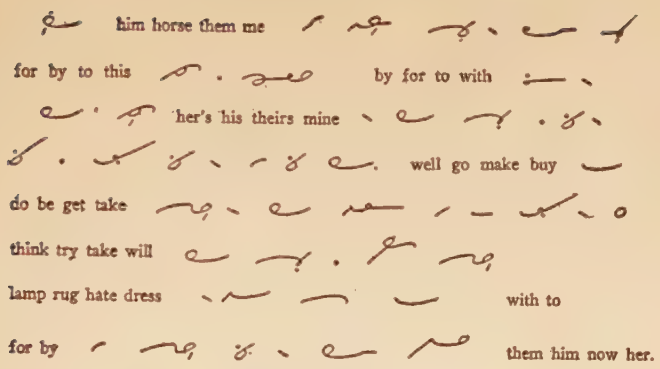
Accuracy and quantity completed will be counted.

Time: 5 minutes Score.....

TEST I-D

Reading Ability





BOOKKEEPING

Another test in the Commercial field that deserves consideration is the Carlson Bookkeeping Tests. These tests are primarily based on the text "Twentieth Century Bookkeeping," but appear to be well adapted to use with other texts. Series *A* has now 9 tests developed, Series *B* 2 tests. These tests are as yet not all well standardized, although they are in the process of standardization. Each one tests somewhat different phases of bookkeeping. The phases of bookkeeping measured in Test No. 1 are:

Section 1, Journal Entries.
Section 2, Ledger Entries.
Section 3, Statements.
Section 4, Sentences.

Test No. III includes:

Section 1, Special Journals.
Section 2, Journalizing.
Section 3, Posting and Trial Balance.
Section 4, Financial Statements.
Section 5, Closing the Ledger.
Section 6, Theory.

Test No. V includes:

Section 1, Recording Transactions.
Section 2, Posting and Trial Balance.
Section 3, Correcting and Adjusting Entries.
Section 4, Classification of Accounts.
Section 5, Closing and Post-Closing Entries.

Test No. VI includes:

Section 1, Adjusting Entries.
Section 2, Trial Balance.
Section 3, Classification of Accounts.
Section 4, Closing and Post-Closing Entries.
Section 5, Theory and General Principles.

The following statement from a very careful teacher of bookkeeping may give an estimate of the value of these tests.

The diagnostic value of these tests makes them highly commendable. They give a good measure of the ability of the pupils and at the same time reveal the weak points for remedial measures. The content of the tests is in accord with what we are teaching. The individual scores correlate high with the class standing of the pupil. Equivalent and duplicate forms would add to the value of these tests.

COMMERCIAL ARITHMETIC

The Kinney Commercial Arithmetic Scales are intended to test the pupil's ability to solve problems similar to the problems found in the business world. The test is used by high school pupils taking Commercial Arithmetic.

Test *A* is intended for use at the end of the first semester. This test covers work in the computation of fundamentals, especially work in short-cut methods. Test *B* presents fairly simple problems in commercial arithmetic and is intended for use at the end of the first semester. Test *C* is devised for use

at the end of the second semester. It involves the use of problems to help the child diagnose his difficulties. Forms 1 and 2 are available. The test is standardized.[6]

COMMERCIAL TESTS

Business Aptitude Test. By F. A. Moss, K. T. Omwake, T. Hunt. Published: 1926. Purpose: to assist in giving an indication of one's natural adaptability to various types of business vocations of junior grade. Range: junior and senior high school. Time to give: 35 minutes. Time to score: 4 minutes. Standardized: for business school and employment groups.
Publisher: Center for Psychological Service, Washington, D. C.
Price: $7.50 per 100, complete.

Clerical Tests. By Bureau of Public Personnel Administration. Published: 1926. Purpose: to measure clerical ability. Range: applicants for clerical positions. Time to give: 90 minutes.
Publisher: Bureau of Public Personnel Administration.

Cody's Commercial Tests and How to Use Them. By Sherwin Cody. Published: 1919. Purpose: to measure business ability. Range: no information. Time to give: 5 to 20 minutes. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.20 per copy, postpaid.

Ruggles Test of Clerical Aptitude. By Allan M. Ruggles. Published: 1920. Purpose: to test ability to do clerical work. Range: commercial department of high schools. Time to give: 33 minutes. Time to score: about 5 minutes. Standardized.
Publisher: Bureau of Educational Research, University of Oklahoma.
Price: $1 per package of 25.

Thurstone's Clerical Examination. By L. L. Thurstone. Published: 1919; revised, 1922. Purpose: to measure ability to do clerical work. Range: applicants for clerical positions. Time to give: no time limit. Time to score: short. Standardized.
Publisher: World Book Co.
Price: $1.50 per package of 25 examination booklets, complete; specimen set, 15 cents.

[6] H. L. Smith and Wendell W. Wright, Second Revision of Educational Measurements, Bureau of Coöperative Research, Indiana University, November, 1927.

FILING

Filing Test. By the Scott Company. Published: 1926. Purpose: to be used in selecting filing clerks. Range: filing clerks. Time to give: no time limit. Time to score: short.
Publisher: C. H. Stoelting Co.
Price: $5.60 per package of 25 tests; $16.80 per package of 100; specimen set, 56 cents.

SHORTHAND

Blackstone's Stenographic Proficiency Tests. By E. G. Blackstone. Published: 1923. Purpose: to measure speed and accuracy in typewriting. Range: commercial schools or business institutions. Time to give: 3 minutes. Time to score: 5 minutes. Standardized.
Publisher: World Book Co.
Price: $1 per package of 25, each form.

Hoke Measuring Scale for Gregg Shorthand Penmanship. By Elmer R. Hoke. Published: 1922. Purpose: to measure pupils' ability to write shorthand. Range: pupils studying Gregg shorthand. Time to give: 2 to 5 minutes. Time to score: short. Not standardized.
Publisher: Gregg Publishing Co.
Price: 25 cents.

Hoke Measuring Scale for Knowledge of Gregg Shorthand. By Elmer Hoke. Published: 1922. Purpose: to measure pupils' ability to read Gregg shorthand. Range: pupils studying Gregg shorthand. Time to give: short. Not standardized.
Publisher: Gregg Publishing Co.
Price: 25 cents per copy.

Hoke Prognostic Test of Stenographic Ability. By Elmer Hoke. Published: 1922. Purpose: to predict stenographic ability. Range: pupils studying stenography. Time to give: 20 to 30 minutes. Time to score: 4 minutes. Standardized: in process.
Publisher: Gregg Publishing Co.
Price: 1½ cents per copy.

Rollinson Diagnostic Shorthand Tests. By Ethel A. Rollinson. Published: 1924. Purpose: to measure class work and individual work in the first four chapters of *Gregg Shorthand Manual*. Range: students studying Gregg shorthand. Time to give: from 2 to 5 minutes for each test. Time to score: short. Standardized.
Publisher: Gregg Publishing Co.
Price: 4 cents for one copy of each of the four tests; specimen set, 25 cents, complete.

TYPING

Thurstone Proficiency Test for Typists. By L. L. Thurstone. Published: 1920. Purpose: to select office clerks, typists, and junior stenographers. Range: clerks and stenographers. Time to give: no time limit. Time to score: short. Standardized. Publisher: World Book Co.
Price: $1.50 for 25 booklets, complete; specimen set, 15 cents.

BOOKKEEPING

Carlson Bookkeeping Tests. By P. A. Carlson. Published: 1925. Purpose: to test a pupil's accuracy in the principles of bookkeeping. Range: pupils studying bookkeeping. Time to give: no time limit. Time to score: short. Partially standardized. Publisher: South-Western Publishing Co.

BIBLIOGRAPHY

Birath, Stanley B. *A Tentative Standard Progressive Achievement Test in Bookkeeping.* Thesis, Temple University, Philadelphia.

Cook, William A. "An Effort to Measure Typing Efficiency," *Journal of Educational Research,* Vol. 2: pp. 49-59, January, 1925.

Hoke, E. R. "The Improvement of Speed and Accuracy in Typewriting," *The Johns Hopkins University Studies in Education No. 8,* The Johns Hopkins Press, Baltimore, 1922.

Vavra, Minnie A. "Success in Typewriting," *Journal of Educational Psychology,* Vol. 16: pp. 487-92, October, 1925.

CHAPTER XVII

Health and Physical Education

Health and physical education apparently have an important place in the minds of the curriculum-makers for our public schools, because an increasing emphasis has been given to this subject during the last ten years. For some time physiology and hygiene have had a place in the elementary course of study. More recently health education, physical education, and play are being given a prominent place.

In the earlier stages of development much emphasis was given to knowledge of physiology and anatomy and to drill in formal gymnastics. During more recent years the emphasis has shifted to knowledge of health habits, the practice of health habits, and the creation of proper attitudes toward healthful living. In the physical training work much is being done to develop the physical body through the kind of play and the games that are of greatest benefit in increasing vitality as well as in giving enjoyment. The work of the school through its medical examination and inspection departments has brought forcefully to the attention of teachers, supervisors, and administrators the need for attention to health and physical fitness. Approximately 8 per cent of the total time allotment in the first six grades of the elementary schools is given to Health Education and Physical Training.[1] This per cent does not include the recess periods in which there is much opportunity for play. The addition of the time of recess periods would approximately double the per cent of time allotment for these subjects.

The objectives of the teaching of Health and Physical Train-

[1] *Second Yearbook,* Department of Superintendence of N. E. A. (1924), p. 144.

ing are not generally agreed upon. The following statements will give some idea of those that are commonly given as applicable to elementary school and high school.

Ability to control one's dietary in such ways as to make one's food contribute in maximum measure to one's physical well-being.

Ability to keep the body mechanism properly oxygenated.

Ability to utilize muscular exercise as a lifelong means of maintaining a high level of physical vitality.

Ability and disposition throughout life to engage with pleasure and profit in a varied repertory of games, sports, athletics, such as swimming, skating, hiking, rowing, riding, tennis, golf, ball games of various kinds, running games, dancing, fishing, hunting, canoeing, motoring, camping, athletic events, etc.

Ability to carry one's self and to move and act with ease, grace, and precision.

Ability to maintain postures conducive to the best physical functioning.

Ability to make one's sleep contribute in maximum measure to the development and maintenance of a high level of physical vitality.

Ability to relax physically and mentally at proper times and in proper ways.

Ability to protect one's self from micro-organisms; and to deal with them and their products effectively in case of attack.

Ability to take proper precautions against the spread of disease.

Ability to dress in ways that promote the physical well-being to a maximum degree.

Ability and disposition to maintain personal cleanliness.

Ability to secure that variety or diversity of physical experiences necessary for maximum well-being.

Ability to draw up an individual program of work, play, rest, sleep, meals, etc., best suited to one's physical nature and capacity.

Ability to avoid preventable accidents.

Ability to deal with conditions produced by many kinds of common accidents.

Ability to care for the teeth, eyes, nose, ears, throat, skin, hair and scalp, nails, and feet.

Ability to keep reasonably well-informed, in the degree to be expected of the layman, as to the discoveries of science in the fields of health conservation and promotion.

Ability to take the protective, precautionary, or remedial steps necessary to protect one's self or family from common ailments.

Ability wisely to utilize the services of physicians, nurses, dentists, and other specialists in health and physical upbuilding and maintenance.

Ability to make one's various mental and emotional states and activities contribute in maximum degree to one's physical functioning.[2]

These general objectives do not give an idea of the specific content of courses of study, but they do give general statements of legitimately expected outcomes of education, whether in the home or in the school. How much the school is supposed to contribute toward the attainment of these objectives is not definitely settled. While much progress has been made in the field of Health and Physical Education in both the elementary and high school, there is relatively little scientific evidence of the amount or type of progress that children make as a result of the teaching of these subjects in the schools. This lack is partly a result of failure on the part of those who develop the courses of study for Health and Physical Education to specify definitely the attainments expected, which would provide a basis for accurate measurements of achievement.

Measurement in this field offers many difficulties not easily overcome. Two of the things most desired in health instruction are the development of habits of health and the proper attitudes toward healthful living. These habits and attitudes are difficult to measure. Another obstacle to measurement in this field is the fact that our fund of knowledge as to the best ways of healthful living is constantly changing. Besides this,

[2] Franklin Bobbitt, *How to Make a Curriculum,* Houghton Mifflin Co., 1924.

in the field of physical training it is difficult to find a very accurate measurable index of physical development. In addition to this, if good measures were available, it would still be exceedingly difficult to determine the influence of school instruction in this subject because much of both health and physical development is modified to a marked degree by what takes place outside the control of the school. What takes place outside the school may help or may hinder the work of the school. While the same modifying effects are at work in relation to any of the subjects of the curriculum, they do not influence the result so materially in these other subjects as they do in health and physical training. There is a real opportunity for development of measures of achievement in the lines of work undertaken by the schools in Health and Physical Training. This can be done only by bringing together the parent, the teacher, the nurse, the doctor, the mental hygienist, the physical education supervisor, the school administrator, and the educator trained in the technique of educational measurements, so that all will understand each other's point of view and work toward a common end.

The work should center around four objectives. The first of these is the development of a course of study and activities with specific grade objectives and statements of expected attainments and pupil activities. The second is the development of methods of administration of the various special activities of the doctor, the nurse, the teacher, the physical training supervisor, etc., so that they may best contribute to the objectives to be attained. The third is the measurement of pupil achievement in the light of these expected attainments. Such measurement would include among other things the measurement of health knowledge, skills, habits, and attitudes with reference to the individual and to the community, measurement of physical development, measurement of physical vitality, measurement of knowledge and proficiency in play, inventory tests, survey tests, diagnostic tests, and classification tests that

would be highly useful for various purposes in the school. The fourth of these objectives is the development of methods of diagnostic and remedial work needed to further the use of the tests.

Three of the various types of measures that have been developed in the field of Health and Physical Education will be briefly discussed. Health tests, anthropometric measures, athletic proficiency tests, physical capacity tests, and other types of measurement in this field may be studied by use of the references at the end of the chapter.

The Gates-Strang Health Knowledge Test affords, in the method of selection of its content, an excellent example of a desirable procedure in test validation. The content is based on a careful analysis of the facts concerning many phases of healthful living, which were found in twenty selected courses in health offered in rural and city schools in America and in fourteen of the most widely used textbooks on health. Seven hundred and forty-four items or facts were found common to three or more of these sources. These facts were then put into the form of a new-type test. The content of the test and its construction were criticized by a group of experts in the field of work covered by the test. The test was then given to several hundred children in grades 3 to 9 inclusive. After an analysis of the results of this test, certain items were eliminated, leaving 520 items. These remaining items are distributed as follows over the following phases of health education.[3]

Topic	*Number of Exercises or Items*
Food	98
Disease Prevention	78
Physiology	33
Exercise and Posture	30
Cleanliness	27

[3] *Teachers College Record* (June, 1925), p. 870.

Topic	*Number of Exercises or Items*
Fresh Air and Sunlight	21
Mental Hygiene	18
Care of the Eyes	17
Safety	15
Defects, including Malnutrition	14
Clothing	14
Care of the Teeth	14
Water	13
Industrial Hygiene	12
Values of Health	12
First Aid	11
Elimination of Body Wastes	10
Care of the Hair	10
Rest and Sleep	10
Disposal of Garbage and Waste	10
Child Care	9
Names of Scientists and Miscellaneous Items	8
Alcohol	7
Care of Ears, Nose, and Throat	6
Care of the Feet	6
Patent Medicines and Drugs	5
Tobacco	4
Health Laws	3
Public Health Administration	3
Health Organizations	2

The material included in these thirty groups has been broken up and organized into standardized tests containing sixty-four exercises. Form I of this test is given; other forms are available.

Directions for Giving

1. See that each child has a sharpened pencil.
2. Tell children to sit with arms folded until told what to do.
3. Distribute test books with front page up.
4. Give directions for writing name, age, school, city, and date. Write the date on the board. Ask children not to open booklets until told to begin.
5. Read aloud deliberately the directions on the first page of the test booklet. Pause for the children's responses to the answers of each sample. Be sure each child has marked the exercise correctly.

6. Then say, "Now when I say Ready, turn the page and begin with number one. Mark only the one best answer. When you have finished one page turn to the next. Do not stop until you have finished all. Do not skip any. If you finish before the time is up, look over your answers to be sure they are right. Ready. Begin." Note exact time immediately.
7. If children ask any questions about the meaning of questions just say, "Do the best you can."
8. See that the room is quiet throughout the period.
9. Be ready to supply a pencil to any child who breaks his.
10. When exactly forty minutes have passed, say, "Pencils up," and collect papers immediately.

FORM I[4]

Grades 3 to 12

Name ..... Boy or Girl?..... Date.....

How old are you?.... When is your birthday?.... Grade...

Name of your school..... Town or city.....

Directions: Here are some questions about health. Five answers are given to each question. Read carefully each question and the five answers. Then mark only the *one* best answer. If you do not know which answer is best, mark one anyway. If more than one is marked your answer will be called wrong.

Sample Exercises:

1. We should have fresh air
   .X.all of the time
   ....in the daytime but not at night
   ....at night but not during the daytime
   ....especially in summer
   ....when we begin to get a headache

[4] From Gates-Strang Health Knowledge Test, prepared by A. I. Gates and R. Strang, published by Bureau of Publications, Teachers College. Copyright by Teachers College, 1925.

1. It is best for children to eat
   ...at regular times each day
   ...whenever they want to eat
   ...six or seven times a day
   ...just before going to bed
   ...no breakfast

2. At meal time we should
   ...eat quickly and hurry back to school
   ...take a sandwich and run out to play
   ...eat in a hurry so we can spend more time outdoors
   ...sit down at the table and eat slowly
   ...play with the knife and fork

3. Cooked meat or vegetables should be kept
   ...in cool, damp cellars
   ...in the sunlight
   ...in the air, uncovered
   ...in a cold place, covered
   ...in a warm, dry closet

4. Alice is thin and pale. Besides rest and some outdoor exercise she should probably have
   ...candy between meals
   ...medicine
   ...meat three times a day
   ...more milk, bread, butter, spinach
   ...more fried potatoes with gravy

5. Flies should be kept away from food because they carry
   ...soot
   ...odors
   ...germs
   ...pollen
   ...dust

6. There is less danger of your giving a cold to some one else if you
   ...cover your face when coughing and sneezing
   ...use same glass to drink from
   ...lend him your handkerchief
   ...exchange pencils with him
   ...sit close to him

7. A person who has trouble with his eyes should
   ...buy glasses at 10-cent store
   ...borrow glasses from some one in the family
   ...ask a neighbor what kind of glasses to get
   ...have glasses fitted to his eyes by an eye doctor
   ...not do anything about it

8. Of these foods the best to choose for breakfast is
   ...buns
   ...hot biscuit
   ...pancakes
   ...pie
   ...hot cereal

9. Boys and girls who wish to be healthy should choose for a lunch at the school lunchroom
   ...caramel candy, hot biscuit
   ...cherry pie, doughnuts
   ...nut cake and maple ice cream
   ...roast pork and apple sauce
   ...hot soup, baked potato, milk

10. One reason why people should not smoke is that it
    ...makes the skin yellow
    ...makes people stupid
    ...causes chills and fever
    ...often harms the lining of nose and throat
    ...is hard to learn how to smoke

11. In washing and drying dishes
    ...clean towels are needed once a week
    ...cold water is best
    ...dishes should not be rinsed because it takes too much time

...very hot water and soap are needed every time
...soap is bad because the dishes may taste soapy afterwards

12. The facts about keeping well which we learn now
...will always be true
...may change as new discoveries are made
...will certainly be the same fifty years from now
...need never be thought of again
...are just rules to be memorized

13. One of the chief causes of decayed teeth is
...having food lacking in some elements while the teeth are growing
...having toothaches
...using our teeth too little for biting threads and hard candy
...using a toothbrush too much
...not enough meat in the diet

14. Safety in games means
...never thinking of danger
...tripping others for fun
...shooting anywhere with air guns or bean shooters
...paying no attention to where you throw balls or swing bats
...watching out so as not to trip or hit other people

15. Of the following, the best way to spend the time right after school is to
...study tomorrow's lessons
...play in the attic
...play outdoors in the sunshine
...play in the barn or woodshed
...stay indoors and read a book

16. The best way to sleep is
...with heavy covers in a room without windows
...with light, warm covers, and windows open top and bottom
...in heated room with windows open a crack
...in a cold room with windows closed
...with covers pulled over the head

17. For eating a good noonday meal, school children should plan to use about
...5 minutes
...10 minutes
...15 minutes
...½ hour
...1½ hour

18. We should visit a dentist
...only when we have a toothache
...every two or three years
...regularly twice a year
...only if our teeth are crooked
...only to have teeth pulled

19. Long and regular hours of sleep
...help keep children well
...cause nervousness and illness
...are necessary only for babies
...make children sleepy in the daytime
...keep children from doing good school work

20. A growing boy should eat roast beef or other meat
...more often than he drinks milk
...not more than once a day
...twice a day

...more often than he eats potatoes
...three times a day

21. If you should get a low mark in school work, it will help matters to
...decide you are no good, and stop working
...ask the teacher's advice and follow it
...cry, and tear up your report card
...tell your parents the teacher has it in for you
...pretend you are sick and can't go to school

22. We help to keep from having accidents on the water when
...we move about as we like in canoes or rowboats
...we hold people under the water for fun
...we try to learn to swim in deep water
...we "take dares" in diving and jumping into the water from high places
...we learn how to swim and to save a drowning person

23. If the organs of our bodies could talk they would say
..."wear tight belts"
..."wear garters"
..."wear high-heeled shoes"
..."wear loose clothes"
..."wear high collars"

24. One reason why green leafy vegetables are healthful is that they contain
...protein and sodium
...starch and sugar
...iron and vitamins
...fat and potassium
...carbohydrates and water

25. In choosing what clothes to put on when you dress in the morning you should think whether or not they
...are suited to the weather
...are new or old
...are stylish
...are better than other people's
...will show the dirt

26. Of these five foods, the most important one for children is
...meat
...butter
...fish
...sugar
...milk

27. Every one should start killing flies
...in July when there are a great many
...as soon as any cases of typhoid fever are reported
...as soon as the first fly appears in the spring
...in mid-summer
...in the fall when there are not so many

28. Of the following the most healthful drink for children is
...ginger ale
...tea
...coffee
...soda water
...orange juice

29. Of the following the best reason why patent medicines should be avoided is
...they are more expensive than other medicines
...they often contain drugs which are bad for us
...they make people feel lively
...doctors do not use them
...they make people poor

30. We get the best air to breathe
...from a hot-air furnace
...from electric fans

...by being outdoors in the sunshine
...by opening the doors into other rooms
...by keeping the night air out

31. Of the following the most healthful meal for a growing girl is one of
...chicken pie
...string beans, asparagus, cabbage
...milk, roast beef, boiled eggs
...milk, bread and butter, spinach, dates
...roast beef, potatoes, turnips, beets, chicken salad, jelly, mince pie, milk

32. One way tuberculosis is spread is by
...mosquitoes and rats
...our coughing and spitting
...our going to a sanitarium
...high winds
...fish and birds

33. While eating, it is a good thing to
...talk about your troubles
...worry about what you are eating
...study your lessons
...say nothing to anybody
...feel comfortable and happy

34. A good way to learn to sit well is
...to wear shoulder braces
...not to take active exercise
...to sit reading when not at school
...to look at the ground all of the time
...to use chairs and desks at school of just the right size and shape

35. When a child gets a cinder in his eye one good thing to do is to
...rub his eye back and forth with his hand
...not wash his eye
...have some one, who knows how, get the cinder out
...wipe his eyes with his handkerchief
...leave it alone until it comes out itself

36. Vitamins are especially necessary for
...furnishing energy
...giving flavor to food
...increasing health and growth
...regulating body temperature
...preventing typhoid fever

37. Iron is found in largest amounts in
...white bread and butter
...green vegetables and egg yolk
...ice cream and cake
...cheese and crackers
...strawberries and cream

38. Constantly breathing through the mouth is harmful because
...it does not look well
...it is always a sign of diseased tonsils
...it is rude
...it annoys other people
...it carries cold uncleaned air into the lungs

39. On a farm, manure should be
...kept in covered bins
...left in barnyard to be gradually washed away by rain
...used to make bricks
...left in uncovered piles until needed
...thrown into the river

40. When studying, we should have
...the sunshine on our book
...a bright light on the table in front of us
...a lamp or electric light over our shoulder so that our book is in our own shadow

...shaded light from above and behind us, without shadows on our book
...both daylight and lamplight

41. Every day a child should drink
...about one cup of water
...four to eight cups of water
...almost two cups of water
...at least three quarts of water
...at least four quarts of water

42. One reason why milk is important is because
...it aids growth
...the government inspects it
...it is the only food a school boy needs
...it prevents typhoid fever
...it doesn't have to be chewed

43. One reason why food should be well chewed is because
...it can be swallowed more quickly
...the saliva begins to digest starchy food
...less food is needed
...animals chew their food
...it is polite to chew thoroughly

44. One good simple rule for preventing constipation is to
...take castor oil once a week
...take a sugar-coated pill every night
...drink one or two glasses of water on rising in the morning
...eat no raw fruit
...eat meat at least twice a day

45. A good place for mosquitoes to be born is in
...tin cans half full of water
...tin cans full of dirt
...tin cans with food left in them
...dry garbage cans
...ash cans

46. The players on champion athletic teams do not touch drinks with alcohol in them because
...alcohol tastes bitter
...alcohol decreases accuracy and quickness of action
...alcohol is expensive
...alcohol makes one fight
...alcohol makes one too hot

47. In general, the best exercise for fourth grade boys and girls is
...work on the farm
...active outdoor play
...running errands
...exercise in the gymnasium
...five minutes of arm exercises every day

48. If no germless bandages are at hand for tying up a cut the best thing to do is
...to use the handkerchief from your pocket
...to use a piece of your skirt
...to use court plaster or adhesive tape over the cut
...to use a boiled piece of cloth that is allowed to touch nothing but the cut
...to hold your hand on the cut until you get a bandage

49. Lime is needed in largest amounts in building
...bones
...skin
...muscle
...hair
...nerves

50. A child in the sixth grade should keep his book
...5 to 8 inches in front of his eyes

...8 to 10 inches in front of his eyes
...14 to 20 inches in front of his eyes
...20 to 25 inches in front of his eyes
...25 to 30 inches in front of his eyes

51. Vaccination for smallpox, typhoid, and other diseases is used to
...clean out the body as castor oil does
...kills germs as a germ killer does
...make new substances grow in the blood that destroy germs
...find out if one is going to get the disease
...help cure the disease after one becomes sick with it

52. In dusting woodwork the housekeeper should use
...a feather duster
...a dry cloth
...a damp or oiled cloth
...a wet cloth
...a whisk broom

53. The best location for a well is
...at the foot of the hill on which the house and barn stand
...ten feet from the outhouses
...half a mile from the house or school
...where it can get surface water so as always to have plenty of water in it
...on much higher ground than the barn or water-closet

54. People who take habit-forming drugs are usually those who are
...brave
...successful
...happy
...efficient
...unsuccessful

55. A good American citizen
...buys and drinks no alcoholic beverages
...acts as if prohibition were a joke
...thinks it is clever to break the law regarding alcoholic drinks
...sells alcoholic drinks when he gets a chance
...is not willing to try a great experiment

56. School children should carry their book-straps or brief cases
...on the strongest arm
...on one hip
...by changing hands every few minutes
...always in the right hand
...always in the left hand

57. The amount of food we ought to eat depends mostly upon
...our appetites
...our size and what we do
...the time we have to eat our meals
...the size of our salary
...the time of year

58. The best source of starch is
...milk
...leafy vegetables
...cereals and bread
...eggs
...sugar

59. The average amount which the stomach of a grown person can hold is
...½- 1 pint
... 2- 3 pints
... 4- 5 pints
... 6- 8 pints
... 9-10 pints

60. The largest amount of protein is to be found in

...rice, oatmeal, wholewheat bread
...milk, fish, eggs
...carrots, potatoes, turnips
...apples, bananas, oranges
...candy, cake, ice cream

61. The best temperature for a room in which one is sitting reading is between
...50 and 54 degrees
...65 and 68 degrees
...72 and 75 degrees
...76 and 80 degrees
...80 and 85 degrees

62. A good average number of calories for lunch for an average boy 13 years old is
... 300- 400 calories
... 400- 500 calories
... 500- 600 calories
... 700- 900 calories
...1000-1200 calories

63. The man who did most to establish the germ theory of disease was
...Goldsmith
...Cook
...Pasteur
...Carpentier
...Harvey

64. The average number of calories in a half-pint glass of whole milk is
... 50 calories
...150 calories
...100 calories
...250 calories
... 75 calories

TABLE 48.—AVERAGE SCORES OBTAINED IN A TYPICAL NEW YORK SCHOOL ON GATES-STRANG HEALTH KNOWLEDGE TEST

At End of	*Actual Mean Score*	*Smoothed Curve Score*	*Approximate Number of Pupils*
3A	15	15	28
3B	..	17	..
4A	..	20	..
4B	21	23	66
5A	..	26	..
5B	31	30	64
6A	35	33	53
6B	36	36	57
7A	41	39	34
7B	42	42	31
8A	44	45	32
8B	..	48	..

While it is recognized that health knowledge is not the most important outcome to be expected from health instruction, nevertheless it is basic to rational motives, the formation of proper habits, and the development of desired attitudes. The

520 items given in the monograph [5] describing this work constitute an excellent basis for inventory purposes and for diagnostic testing of health knowledge. The facts that have been taught during the preceding semester as well as those that are to be taught during the semester in question might well be tested to determine the status of the class on these phases of health knowledge. By selecting the desired items from the list of 520, an inventory test for this purpose could easily be devised. In the same manner diagnostic tests could be developed. By the use of diagnostic charts like those described in previous chapters, which would show both pupil and class difficulty on various items tested, it would be possible to determine the phases of the work that need reteaching and the pupils that need special help.

An Analysis of Instruction for Habits and Practices in Health and Accident Prevention, by E. George Payne,[6] is a rating scale that can be of use in developing standards and emphasizing vital phases of health and accident work.

Much work has been done in the standardization of a height-weight-age table by Thomas D. Wood and B. T. Baldwin. These tables are to be found in a great many schoolrooms. One of the best and most easily obtained measures of general health is the height-weight index.

Two tables, one for boys and another for girls, are given.

After the measures of height in inches and weight of the pupil are accurately taken, the interpretation of these facts can be obtained by finding in the table the column at the top that represents the age of the child that has been measured. Follow down this column until even with the horizontal line which represents the child's height. The figures that indicate height are at the extreme right and left of the table. The figure at that place in the column that cuts the horizontal line

[5] Gates-Strang Health Knowledge Test, Bureau of Publications, Teachers College.

[6] Published as a scale by Public School Publishing Co., also by the author in "Education in Health," Lyons and Carnahan, 1921.

TABLE 49.—WEIGHT-HEIGHT-AGE TABLE FOR GIRLS OF SCHOOL AGE[7]

Height (inches)	Average weight for height (lbs.)	5 Years	6 Years	7 Years	8 Years	9 Years	10 Years	11 Years	12 Years	13 Years	14 Years	15 Years	16 Years	17 Years	18 Years	Height (inches)
38	33	33	33													38
39	34	34	34													39
40	36	36	36	36*												40
41	37	37	37	37*												41
42	39	39	39	39*												42
43	41	41	41	41	41*											43
44	42	42	42	42	42*											44
45	45	45	45	45	45	45*										45
46	47	47*	47	47	48	48*										46
47	50	49*	50	50	50	50	50*									47
48	52		52	52	52	52	53*	53*								48
49	55		54	54	55	55	56	56*								49
50	58		56*	56	57	58	59	61	62*							50
51	61			59	60	61	61	63	65							51
52	64			63*	64	64	64	65	67							52
53	68			66*	67	67	68	68	69	71*						53
54	71				69	70	70	71	71	73*						54
55	75				72*	74	74	74	75	77	78*					55
56	79					76	78	78	79	81	83*					56
57	84					80*	82	82	82	84	88	92*				57
58	89						84	86	86	88	93	96*	101*			58
59	95						87	90	90	92	96	100	103*	104*		59
60	101						91*	95	95	97	101	105	108	109	111*	60
61	108							99	100	101	105	108	112	113	116	61
62	114							104*	105	106	109	113	115	117	118	62
63	118								110	110	112	116	117	119	120	63
64	121								114*	115	117	119	120	122	123	64
65	125								118*	120	121	122	123	125	126	65
66	129									124	124	125	128	129	130	66
67	133									128*	130	131	133	133	135	67
68	138									131*	133	135	136	138	138	68
69	142										135*	137*	138*	140*	142*	69
70	144										136*	138*	140*	142*	144*	70
71	145										138*	140*	142*	144*	145*	71
Age—years			6	7	8	9	10	11	12	13	14	15	16	17	18	
Average height (inches)	Short		43	45	47	49	50	52	54	57	59	60	61	61	61	
	Medium		45	47	50	52	54	56	58	60	62	63	64	64	64	
	Tall		47	50	53	55	57	59	62	64	66	66	67	67	67	
Average annual gain (lbs.)	Short		4	4	4	5	6	6	10	13	10	7	2	1		
	Medium		5	5	6	7	8	10	13	10	6	4	3	1		
	Tall		6	8	8	9	11	13	9	8	4	4	1	1		

7 Prepared by Bird T. Baldwin, Ph.D., Iowa Child Welfare Research Station, State University of Iowa, and Thomas D. Wood, M.D., Columbia University.
The figures not starred represent exact averages in round numbers.
The starred figures represent smoothed or interpolated values.

TABLE 50.—WEIGHT-HEIGHT-AGE TABLE FOR BOYS OF SCHOOL AGE[8]

Height (inches)	Average weight for height (lbs.)	5 Years	6 Years	7 Years	8 Years	9 Years	10 Years	11 Years	12 Years	13 Years	14 Years	15 Years	16 Years	17 Years	18 Years	19 Years	Height (inches)
38	34	34	34*														38
39	35	35	35														39
40	36	36	36*														40
41	38	38	38	38*													41
42	39	39	39	39*	39*												42
43	41	41	41	41*	41*												43
44	44	44	44	44	44*												44
45	46	46	46	46	46*	46*											45
46	48	47*	48	48	48	48*											46
47	50	49*	50	50	50	50*	50*										47
48	53		52	53	53	53	53*										48
49	55		55	55	55	55	55	55*									49
50	58		57*	58	58	58	58	58*	58*								50
51	61			61	61	61	61	61	61*								51
52	64			63	64	64	64	64	64	64*							52
53	68			66*	67	67	67	67	68	68*							53
54	71				70	70	70	70	71	71	72*						54
55	74				72*	72	73	73	74	74	74*						55
56	78				75*	76	77	77	77	78	78	80*					56
57	82					79*	80	81	81	82	83	83*					57
58	85					83*	84	84	85	85	86	87					58
59	89						87	88	89	89	90	90	90				59
60	94						91*	92	92	93	94	95	96				60
61	99							95	96	97	99	100	103	106*			61
62	104							100*	101	102	103	104	107	111	116*		62
63	111							105*	106	107	108	110	113	118	123	127*	63
64	117								109	111	113	115	117	121	126	130*	64
65	123								114*	117	118	120	122	127	131	134	65
66	129									119	122	125	128	132	136	139	66
67	133									124*	128	130	134	136	139	142	67
68	139										134	134	137	141	143	147	68
69	144										137	139	143	146	149	152	69
70	147										143	144	145	148	151	155	70
71	152										148*	150	151	152	154	159	71
72	157											153	155	156	158	163	72
73	163											157*	160	162	164	167	73
74	169											160*	164	168	170	171	74
Age—years		5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Average height (inches)	Short	43	45	47	49	51	53	54	56	58	60	62	64	65	65		
	Med.	46	48	50	52	54	56	58	60	63	65	67	68	69	69		
	Tall	49	51	53	55	57	59	61	64	67	70	72	72	73	73		
Average annual gain (lbs.)	Short	3	4	5	5	5	4	8	9	11	14	13	7	3			
	Med.	4	5	6	6	6	7	9	11	15	11	8	4	3			
	Tall	5	7	7	7	7	8	12	16	11	9	7	3	4			

8 Age is taken at the nearest birthday; height at the nearest inch; and weight at the nearest pound. A child is considered 6 years old at any time between 5½ and 6½ years.
The figures not starred represent exact averages in round numbers.
The starred figures represent smoothed or interpolated values.

represents the normal weight for the age and height of the pupil measured. Compare the weight of the pupil measured with the weight given in the table. The difference found between these two figures indicates the physical condition of the pupil. If his weight exceeds the standard given, he is marked plus. If his weight is below the standard given, he is marked minus. Exactly what per cent of underweight should be considered serious is not entirely agreed upon. Perhaps from 7 to 10 per cent below standard is most generally accepted as indicating a state of malnutrition.

PHYSICAL AND ATHLETIC PROFICIENCY

Measures of physical and athletic proficiency are usually based on a rather wide range of physical capacity tests and of athletic events in which standards can well be established. An attempt to measure physical ability and to obtain an energy index has been made by Sargent.[9] Sargent's test is an individual one but relatively simple to administer. The height of a jump made by the pupil is tested by measuring how much the top of the head is moved up in the jump. This is a measure of how much the whole body is moved up. This distance is then multiplied by the weight of the pupil, and the product is divided by the height of the pupil. The formula is as follows: $\text{Index} = \frac{\text{weight} \times \text{jump}}{\text{height or stature}}$. Thus, if an individual weighed 120 pounds, jumped 15 inches, and was 64 inches in height, the index would be $\frac{120 \times 15}{64} = 33.75$. Neither the validity nor the reliability of this test has been proved. It suggests a possible short general measure that might be of value.

Rogers[10] has developed physical capacity tests. His tests

[9] "The Physical Test of a Man," *School and Society*, Vol. 13, No. 318, Jan., 1921.

[10] "Physical Capacity Tests in the Administration of Physical Education," Teachers College Contributions to Education, No. 173, New York, 1925.

were selected to measure vital capacity and the strength of the larger muscle groups. The seven tests used by him were:

1. Forced lung capacity, measured with the wet spirometer, in cubic inches.
2 and 3. Strength of grip (of forearm) (flexion) of each hand, measured by grasping the elliptical manometer or hand-dynamometer, in pounds.
4. Strength of back (extension) measured by lifting with knees straight, using the back and leg dynamometer, in pounds.
5. Strength of legs (extension) measured by lifting, carrying the weight on the thighs, with the back and leg dynamometer.
6. Strength of arms (and shoulder girdle) (extension) by "push-ups" from straight-arm hang on the parallel bars, as many times as possible.
7. Strength of arms (and shoulder girdle) (flexion) by "chinning" or "pull-ups," using eight-inch rings attached to an overhead ladder.

The scores made on these tests were combined into a composite measure called a Strength Index.

The validity of these tests was proved by a comparison of the results of the physical capacity tests with tests of outdoor and indoor athletic proficiency. The reliability for the Strength Index is unusually high. A coefficient of correlation of .94 was obtained between scores of the same pupils on the Strength Index taken at periods four months apart. The author says that "all tests should be given each individual within ten to twelve minutes."

Rogers attempted to develop also an Index of Physical Fitness on the assumption that a measure of muscular efficiency (Strength Index) is also a measure of physical fitness. To do this, he found it necessary to correct the Strength Index for each age and weight. This correction gave a measure called a Fitness Index. This index is a measure of the physical capacity of any given age and weight. Thus a boy of twelve

weighing seventy-seven pounds (both of these are normal) whose Strength Index was above the average for that age and weight would also have a Fitness Index above the average. According to Rogers, this would mean that this particular boy is above the average in Physical Fitness.

These two measures will be of value if they prove to be easily applied and interpreted by teachers of physical education. There are two purposes that these measures might well serve. The first of these is classification of pupils for physical education work in the schools. In many schools no attempt is made to classify the pupils for their physical training work. In other schools the crude measures of weight or height or age are used. By the use of the Strength Index, classification can be made on a basis that will give nearly equated groups. The second purpose served is that the use of the Fitness Index makes possible a selection of those pupils who need particular types of physical activities. Rogers says:

A boy 25 per cent above the norm (according to the Fitness Index) does not need the ordinary routine exercise; he should be given special activities for the development of other characteristics, while the boy any distance below the norm should be given special attention; those far below the norm should receive a thorough medical examination.

Athletic Badge Tests for Boys and Girls were developed by the Playground and Recreation Association of America, and published in 1923. The purpose of these tests is to raise the standard of physical efficiency among school children by a standard measure of athletic proficiency. The tests can be used in the elementary and high school.

The Athletic Badge Test for boys is divided into four divisions of increasing difficulty, each including climbing, jumping, running, and throwing tests. The Athletic Badge Test for girls is similarly divided into three divisions, each one including balancing, running, and throwing tests, and also tests based upon games.

TESTS IN HEALTH AND PHYSICAL EDUCATION

Athletic Badge Tests for Boys and Girls. By Playground and Recreation Association of America. Published: 1923. Purpose: to raise the standard of physical efficiency among school children. Range: elementary and high school. Time to give: no time limit. Time to score: short. Standardized: no information.
Publisher: Department of the Interior, Bureau of Education, Washington, D. C.
Price: 5 cents per copy.

Gates-Strang Health Knowledge Test, Form I. By Arthur I. Gates and Ruth Strang. Published: 1925. Purpose: to test health knowledge. Range: grades 3-12. Time to give: 40 minutes. Time to score: short. Standardized tentatively.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: $3 per 100.

Rogers Physical Capacity Tests in the Administration of Physical Education. By Frederick Rand Rogers. Published: 1925. Purpose: to measure vital capacity and the strength of large muscle groups. Range: junior and senior high school. Time to give: 5 to 10 minutes. Time to score: short. Standardized.
Publisher: Bureau of Publications, Teachers College, Columbia University.
Price: $1.25 per copy.

BIBLIOGRAPHY

Bobbitt, F. "Objectives of Physical Education," *American Physical Education Review,* Vol. 26: pp. 229-34, May, 1921.

Brace, David K. *Measuring Motor Ability,* A. S. Barnes & Co., 1927.

Englehardt, J. L. "A Test of Physical Efficiency," *Journal of Educational Psychology,* Vol. 15: pp. 573-98, December, 1924.

Foster, W. C. "Test of Physical Efficiency," *American Physical Education Review,* Vol. 19: pp. 632-36, November, 1914.

Gates, Arthur I. and Strang, Ruth. "A Test in Health Knowledge," *Teachers College Record,* Vol. 26: pp. 867-80, June, 1925.

Jones, I. A. "Physical Training Tests," *American Physical Education Review,* Vol. R 30: p. 236, April, 1925.

Pressey, L. C. and Stephens, W. "A Sports Information Test: With Some Evidence Regarding the Close Relation Between

Each Sport and Academic Success," *American Physical Education Review,* Vol. 31 : pp. 755-58, April, 1926.

Rogers, F. R. *Physical Capacity Tests in the Administration of Physical Education,* Teachers College Contributions to Education, No. 173, 93 pp., New York, 1925.

Sargent, L. W. "Some Observations on the Sargent Test of Neuro-Muscular Efficiency," *American Physical Education Review,* Vol. 29: pp. 47-59, February, 1924.

CHAPTER XVIII

Practice Tests

By using much of the technique that has been developed in the making of standard tests, it has already been possible in some subjects to work out standardized practice tests, drill books, and work-books that are being widely used. Practice tests and work-books in a wider range of subjects will doubtless become available soon. While these practice tests differ from each other in many respects, they nevertheless have some common purposes. The following stated purposes, which overlap to some degree, may be considered valid.

(1) To provide for inventory testing at the beginning of the term's work or at the beginning of the study of a particular topic.
(2) To provide standard measures of progress from time to time.
(3) To diagnose and point out pupil difficulties, both special and general.
(4) To diagnose class difficulties and to point out class weaknesses.
(5) To point out weaknesses in the course of study.
(6) To point out weaknesses of methods of instruction.
(7) To supplement the work of the textbook with additional and new material of work type.
(8) To provide practice and drill on important and essential phases of the class work.
(9) To give recurring practice on material learned—in so doing to prevent forgetting.
(10) To provide specific drill and practice to follow remedial work or to provide drill on those phases of the work with which an individual pupil has difficulty.

(11) To save time and labor for the pupil and the teacher.
(12) To provide for self-measurement by the pupil.
(13) To motivate and make more effective learning by the pupil.

In the use of practice tests or drill material, some difficulties have been encountered that will need to be overcome in the future. (1) If practice material is to be separate from the textbook but intended to provide supplemental practice and drill on important phases brought out in the textbook and the course of study, then the drill material should be based directly upon the course of study and on the textbook used. (2) The emphasis in many existing practice tests and drill exercises is entirely too much on the factual side of the subject. Thus practice is afforded on only a few of the phases of the subject, and measurement is made of only a part of the legitimately expected outcomes of instruction in the particular subject. (3) If practice material is to serve the purposes of saving time and of producing more effective learning, the achievement of these purposes must ultimately be proved by scientific evidence rather than, as at present, by mere belief. There is great need for further scientific evaluation of the use of the practice tests and drill material. Many drill pads, practice tests, and workbooks are now being used. To give some idea of how they are used, some description of a selected number of them is given here.

The Clapp Drill Books in Arithmetic [1] are based on a scientific study of pupils' difficulties in Arithmetic.[2] There are five different drill books for use in the five grades—4th, 5th, 6th, 7th, 8th. The material of the various books provides for frequent short drill on phases of the work in which the pupil is

[1] *The Clapp Drill Books in Arithmetic for Grades 4, 5, 6, 7, 8.* Silver, Burdett and Company, 1926.

[2] Frank L. Clapp, "The Number Combinations; Their Relative Difficulty and the Frequency of Their Appearance in Textbooks," Bulletin No. 2 of the Bureau of Educational Research of the University of Wisconsin, July, 1924.

deficient. The various tests and drills have standard scores. Suggestions as to the methods and material to be used in diagnosing the work of the pupil are given.

Section III of the drill books contains test and drill exercises in Addition, Subtraction, Multiplication, and Division. Each of the twelve different exercises in multiplication contains approximately one hundred combinations. A pupil, in taking the test-drills, meets each of the one hundred combinations twice in the first two exercises and twice in each pair of the successive exercises. In covering the twelve test-drills, the pupil meets each of the one hundred multiplication combinations twelve times.

As an example of the way to determine the standing of the class in subtraction by using an exercise as a test, the directions for giving and scoring, interpreting scores, and making use of the results are given here. A test exercise in multiplication is also reproduced.

SECTION III

General Drill Exercises[3]

To the Pupil:

In this section, when you do the first three exercises in Addition you will meet each of the one hundred combinations six times. Likewise, when you do the next three exercises you will meet each of them six times again. The same is true of the last three exercises.

In Subtraction and in Multiplication you will meet each of the combinations twice when you do the first two exercises, and so on with each remaining two.

In Division the exercises as a whole contain each of the combinations four times.

Since this is true you will be sure to get practice on each of the combinations.

When you do an exercise, record your time and the number of examples you have wrong. Then to find your score use the figure on page 15. You can find in which quarter your score

[3] From *The Clapp Drill Book in Arithmetic for the Seventh Year,* by Frank L. Clapp, published by Silver, Burdett and Company.

falls by noticing the numbers above each exercise as explained on page 15.

You should keep a permanent record of your scores on these exercises. Your teacher will explain how this record may be kept.

You can improve your score by simply repeating these exercises, but if you will do the work in Sections I and II, as your teacher directs, your improvement will be more rapid.

Subtraction

Exercise 76

76		87	97	
(1)	(2)	(3)	(4)	(5)
8594	77175	52646	25036	8528
7163	1189	7339	9565	7131
(6)	(7)	(8)	(9)	(10)
71621	4495	9814	9600	6079
9503	1507	2914	7233	3501
(11)	(12)	(13)	(14)	(15)
23500	46707	60028	61547	34962
8184	9266	4182	2644	7635
(16)	(17)	(18)	(19)	(20)
4199	16194	13947	7856	41322
2294	7837	9509	2180	6186
(21)	(22)	(23)	(24)	(25)
28452	8391	98550	64352	6033
9961	2484	9069	5500	2124

To use this page as a test in subtraction, the following steps are taken.

(1) Have pencils sharpened and paper ready on which to write answers. In using this drill and test book, the answer

to each problem is written not on the book but upon a sheet of paper placed just below the problem. One edge of the paper can be used for the first row of problems, another edge for another row, and so on. The paper can be folded if it needs to be.

(2) The teacher says: "Open your books to Exercise 76, page 64. This is a test in subtraction. You are to work the problems and write only the answers on the sheet of paper that you use. (Pupils already know how to place the paper.) Work as rapidly as you can and get the problems right. When you have finished, look at the board and copy down the smallest figure you see there. Do not change the paper in any way after the time is recorded."

(3) The teacher should write on the board the figures 1 to 10 in a vertical line. She should then note the exact time at which the pupils begin work. At the end of the one minute the teacher should erase the 1. At the end of two minutes the teacher should erase the figure 2, and so on to the end of the ten minutes of test time. As each pupil finishes, he copies on his paper the smallest number yet remaining on the blackboard. In this way is recorded the total time required by each pupil to work the exercise.

(4) When pupils have finished, papers are exchanged and scored for number correct according to the answers in the back of the book.

(5) To obtain the score Table 51 should be used.

To illustrate how, from the table given, to find the score that a pupil should have on the test, suppose pupil A did the test in 3 minutes and got two examples wrong. His score is 78. This is found by looking down the column at the top under 2 to the horizontal column marked 3 (time in minutes). The score where the two columns meet is 78. The pupils can themselves easily carry out this method of scoring.

(6) The resulting scores can be compared to the standard given at the top of the test drill sheet (see Exercise 76). The

TABLE 51.—TABLE FOR SCORING CLAPP SUBTRACTION TEST

Number of Examples Wrong

Time in minutes	9	8	7	6	5	4	3	2	1	0
1	10	20	30	40	50	60	70	80	90	100
2	9	19	29	39	49	59	69	79	89	99
3	8	18	28	38	48	58	68	78	88	98
4	7	17	27	37	47	57	67	77	87	97
5	6	16	26	36	46	56	66	76	86	96
6	5	15	25	35	45	55	65	75	85	95
7	4	14	24	34	44	54	64	74	84	94
8	3	13	23	33	43	53	63	73	83	93
9	2	12	22	32	42	52	62	72	82	92
10	1	11	21	31	41	51	61	71	81	91

numbers in heavy type represent the 25 percentile, the median, and the 75 percentile of standard scores. In the example given, a score of 78 was made by the pupil. Comparison of the score with the standards at the top of the page shows that on the average approximately 75 per cent of the pupils in the seventh grade do better work than this pupil, and 25 per cent do poorer work. Since the pupil is below standard, it might be well to find out what are his difficulties. This can be done by diagnosis. His difficulty in subtraction may be due to the fact that he does not know well enough his subtraction combinations when presented singly or when presented in examples with more than single digits. To determine the pupil's difficulty, the teacher can further test the pupil or pupils who are in need of such diagnosis.

Test *A*, Subtraction, can be used to determine the pupils' difficulties in subtraction when the numbers are presented

singly. In Test *A* each of the one hundred subtraction combinations is presented singly. If in the use of this test it is revealed that there are certain subtraction combinations that the pupil taking it does not know, the pupil's difficulty has been at least partially located. After the difficulties have been located both by the teacher and by the pupil, the next step is reteaching.

If the teacher desires to determine what difficulty the pupil might have in subtraction in problems of more than one digit, Test *B* should be given. If the teacher desires to find out just where in such problems the pupil makes a mistake, she can have the pupil work out a problem orally. Take as an example problem 5 in Test *B*: 7052 − 4127 = 2925. While the pupil goes through the problem, the teacher checks on the teacher's page the items in the problem. The problem appears on the teacher's page in the following form.

7052	12 — 7
4127	4 — 2
———	10 — 1
2925	6 — 4

Check in space below for each error of the following types:

1. Reversed digits
2. Forgot borrowing
3. Assumed borrowing

By following the steps taken by the pupil, any error can be found.

A special oral drill device is suggested as a means of reteaching those phases of computation not known. However, as has often been said in other discussions, reteaching differs little from first teaching except in emphasis. If it is desired to drill on certain combinations after reteaching, it is possible by use of the teacher's manual to determine just what part of the drill pad to use. For example, if a pupil needs drill in the multiplication of sixes, sevens, and nines, the teacher by use

of her manual can find that exercises 30 and 31 give drill in these numbers especially. By the use of the manual, drill can be given on those combinations where it is most needed.

If the teacher desires to give drill in multiplication or in any other particular phase of arithmetic she may use concrete problems taken from exercises 130 to 157 in the seventh year book.

The pupil's score on various tests in the drill pad can be recorded on a record card furnished for that purpose. The test and drill books by Clapp have certain good characteristics, some of which may be summarized as follows: (1) They are based on scientific research. (2) They can be used with almost any type of textbook. (3) They provide for testing before drilling. (4) They provide a technique of diagnosing pupil difficulties. (5) They provide for extra drill on specific phases of arithmetic work both in computation and in concrete problems. (6) They provide drill in mental arithmetic. (7) They have standards. (8) They can be used to motivate arithmetic work through the individual chart for each pupil, which shows progress. (9) The books can be used more than once. This is an advantage in cost.

On the other hand, these practice tests and drill books have some weaknesses. (1) They fail to provide enough test and practice material on all the various phases of arithmetic. While computation, which they stress, is the most important of any of the phases of arithmetic, nevertheless vocabulary, principles, and methods of work are also essential phases of arithmetic and deserve attention in a practice book. (2) They necessitate copying some problems out of the book. There is no good reason for making this requirement in a drill or practice textbook except in so far as practice is needed in copying problems.

Another type of practice test book quite different from the Clapp Drill Book is the Arithmetic Work-Book.[4] This type

[4] Ruch, Knight, Studebaker, and Myers, *Arithmetic Work-Book, Grades 4, 5, 6, 7.* Scott, Foresman & Co., 1925.

of work-book is available for grades 4, 5, 6, 7. As an illustration the Work-Book for grade 6 is here discussed. The purposes of this Work-Book are (1) Review of skills already learned, (2) Systematic re-study of the important ideas of Arithmetic, (3) Provision for independent study by children, (4) Provision for weekly measurement of progress. This book consists of twenty-nine units of work, one of these units to be used each week. The Manual suggests that the best day for using these "work units" is Monday or Friday. The following explanation taken from the Manual will give an understanding of the way in which the "work units" like the one given here are to be used.

THE CLASSROOM USE OF THE WORK-BOOK

6M. *The Detailed Plan of Administration.* Page 3 of this *Work-Book* presents to the pupil the method of handling the Study Notes, Extra Practice, and Drills which is recommended by the authors. This plan is based upon definite and conscious principles, and, except for very good reasons, the teacher is urged to adhere to the plan of administration outlined. The following points should help to make these ideas clear:

1. Each left-hand page and the drill on the facing right-hand page together form a "Work Unit." There are 29 of these units in the book, or approximately one per week for the school year. In general, Monday or Friday is probably the best day for using the *Work-Book,* since it is primarily a review and organizing tool.

2. Each Work Unit consists of three parts: the Study Notes, the Extra Practice, and a Drill.

3. The Study Notes of each Work Unit should be taken up first in the class period. Usually these call for a few minutes of silent reading followed by a brief class discussion.

4. The drill should be taken up as soon as the work with the Study Notes has been finished. The directions should be followed closely and exactly 20 minutes should be allowed between the signal to go and the signal to stop. These signals should be spoken clearly.

5. After the 20 minutes are up, the teacher should read the

correct answers to the class. Ordinarily it is better for the pupils to exchange books for purposes of grading.

6. Each pupil should next determine his rating (or his neighbor's rating) from the Standards printed below each drill and record it at once on the Progress Chart.[5]

7. After the rating has been recorded, the pupils should proceed at once to correct all errors made in the examples or problems of the drill. As far as time permits, class discussion of examples or problems failed by a large number of pupils, should follow next. The difficulties are quickly located by a show of hands on the correctness of each of the twenty examples and problems.

8. The teacher should keep a record in the forms provided with each Work Unit and on the Class Progress Chart as a permanent summary of the accomplishment of the class on each drill. This will be of very great value to the supervisor and to the teacher, year by year, as a measure of their efforts to raise the level of achievement in the school system.

As an illustration, Work Unit No. XVI, which consists of study notes, and Drill 16, is reproduced.

WORK UNIT NO. XVI

Study Notes

47. *Review of the Multiplication of Fractions.* There are five things to remember in working an example in the multiplication of fractions. These five things are:

1. Cancel whenever you can.
2. Multiply the numerators to obtain the numerator of the product.
3. Multiply the denominators to obtain the denominator of the product.
4. Reduce the product to simplest form.
5. When there is a whole number in the example it is used as a numerator.

These five points in multiplying fractions are illustrated below:

POINT 1. Cancel first: $\frac{\cancel{2}}{\cancel{3}} \times \frac{1}{\cancel{7}}$

[5] The rating which is recorded on the Progress Chart is the one earned by the pupil *on his first trial.* The pupils should be encouraged to go back and retake drills for the sake of bettering their ratings on *future* drills. Ratings earned on second, third, or later trials must not, however, be recorded in place of the original record.

POINT 2. Multiply numerators: $\frac{\overset{2}{\cancel{4}}}{\underset{3}{\cancel{9}}} \times \frac{\overset{1}{\cancel{3}}}{\underset{7}{\cancel{14}}} = \frac{2}{}$

POINT 3. Multiply denominators: $\frac{\overset{2}{\cancel{4}}}{\underset{3}{\cancel{9}}} \times \frac{\overset{1}{\cancel{3}}}{\underset{7}{\cancel{14}}} = \frac{2}{21}$

POINT 4. In an example like the following, reduce the product:

$$\tfrac{9}{5} \times \tfrac{4}{5} = \tfrac{36}{25} = 1\tfrac{11}{25}.$$

POINT 5. A whole number is used as a numerator:

$$4 \times \tfrac{1}{5} = \tfrac{4}{5}.$$

48. *The Multiplication of Small Mixed Numbers.* The multiplication of small mixed numbers becomes exactly like the multiplication of fractions if the mixed numbers are first changed to improper fractions. Study the samples below:

$4\frac{1}{2} \times 8.$	Change to $\frac{9}{2} \times 8.$
$3\frac{1}{3} \times 5\frac{1}{2}.$	Change to $\frac{10}{3} \times \frac{11}{2}.$
$6 \times 7\frac{3}{5}.$	Change to $6 \times \frac{38}{5}.$
$6\frac{3}{4} \times \frac{2}{7}.$	Change to $\frac{27}{4} \times \frac{2}{7}.$

Now take Drill 16. Examples 10 and 12 give you a chance to practice the multiplication of fractions.

Extra Practice

49. *Practice on the Division of Decimals.* If you have any doubt about how to handle the decimal point in division, read the *Study Notes* on page 30 before doing the following work.

The answer to every example that follows is 3000, 300, 30, 3, .3, .03, .003, or .0003. Select from this list the correct answer by thinking carefully where the decimal point in the quotient would come and what the quotient would have to be. Write each quotient:

1. $1.75\overline{)525}$	2. $175\overline{)52.5}$	3. $17.5\overline{).525}$	4. $.175\overline{)5.25}$
5. $17.5\overline{)52.5}$	6. $.175\overline{)525}$	7. $17.5\overline{).0525}$	8. $175\overline{).0525}$

Copy the examples below and work in the margin, or on scratch paper.

1. $3.2\overline{)1.612}$	2. $81\overline{).3564}$	3. $96\overline{)393.6}$	4. $.009\overline{)882.}$
5. $.45\overline{).2145}$	6. $8.4\overline{)35.28}$	7. $57\overline{)3648}$	8. $7.6\overline{).7220}$

WORK UNIT NO. XVI

Drill 16

Follow the directions as given on the first three drills. *Try to beat all past records.*

1. $69.9 + 7.8 =$

2. $256006 - 48908$

3. 651×77

4. $162.531 - 90.658$

5. $81011 + 34258 + 10185 + 34259$

6. 18.4×61

7. $9\frac{3}{7} + 7\frac{1}{6} =$

8. $77)6469$

9. $.9630 \times 8.6$

10. $6 \times \frac{7}{30} =$

11. $2\frac{1}{2} \div \frac{3}{5} =$

12. $4\frac{1}{2} \times \frac{1}{2}$

13. $8.4)35.28$

14. $88)5986$

15. If $45 = 100\%$, $500\% = ?$

16. $10\frac{1}{5} - 7\frac{13}{15}$

17. Subtract $3\frac{7}{9}$ from $9\frac{1}{6}$

18.

Width	8 in.
Length	12 in.
Height	15 in.
Volume	?

19. What % of 88 is 77?

20. Areas B + C + D equal how much more than Area E? State your answer in per cents. *Ans.* .....

Standards for Drill 16

Number correct.....	0–4	5–6	7	8–9	10	11	12	13	14	15–19	20
Rating.....	0	1	2	3	4	5	6	7	8	9	10

If your rating does not satisfy you, practice on this drill and the *Extra Practice* in your spare moments.

The Progress Chart used with this Work-Book is shown in Figure 19. On it is shown the progress of a pupil during the year. Such a chart is a valuable part of any practice test, drill pad, or work-book. The ratings on this chart are in

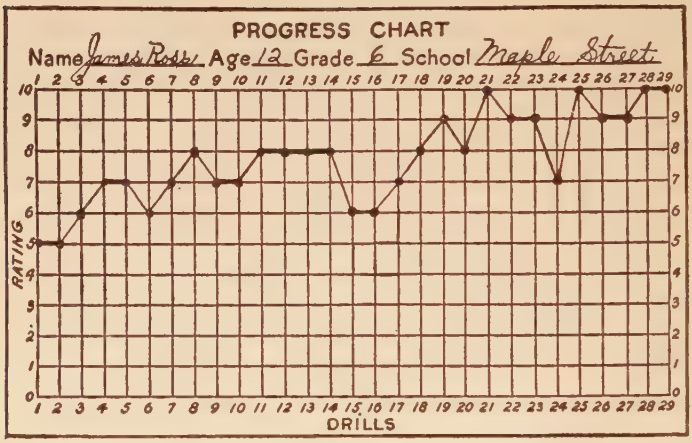


FIG. 19.—An example of the Progress Chart Used with the Work-Book.

terms of standards or norms. A rating of 5 means that an average or median score has been made by the pupil. If a pupil has a rating of 8, it means that he has made a score that equals or exceeds what 80 per cent of sixth grade pupils usually do but that 20 per cent of the sixth grade pupils usually make better scores than he has made on this drill test.

The manual suggests that in making a diagnosis of pupil difficulties the following procedure be used. (1) Determine the number of correct solutions for each example. For instance, if there are thirty pupils in the class and if problem 1 is worked by all correctly, it would be marked 30. Write this number on the record provided for it. Find this same fact for each of the problems. In this manner the number of pupils working each problem correctly is shown on the Teacher's

Diagnostic Record. (2) List for special emphasis those particular problems which were missed by a large number of pupils.[6]

Some of the valuable things about this Work-Book are: (1) It covers most of the important and useful phases of arithmetic study. (2) It has been scientifically developed, and its content is valid. (3) It provides "systematic review" on the ideas, principles, vocabulary, and methods of work as well as on computation. (4) The test and drill exercises are of a mixed type. Problems in addition, decimals, story problems, and so on are found in the same test drill. (5) Extra practice is provided. (6) The standards or norms are put into usable forms by means of a chart. (7) The Work-Book provides self-measurement.

Three serious adverse criticisms may be made of this type of practice test or work-book: (1) This arithmetic Work-Book is not easily used with certain textbooks. It is not so elastic as practice tests should be if they are to be used with various textbooks. If these work-books were fitted to the particular textbook used by the children, the criticism would entirely disappear. Still better perhaps, if these practice tests or work books were included in a textbook to be used by the children, it would be more satisfactory. (2) The diagnostic value of this Work-Book is limited by the lack of diagnostic tests. By the use of the methods suggested on page 434, it is possible to use the Work-Book to find the problems with which the class has difficulty. To find out what the difficulty is and to give specific work and drill in that phase is not so easy with this Work-Book as it should be. The Compass Diagnostic Arithmetic Tests, by the authors of the Work-Book, can be used with the Work-Book to overcome this objection to some degree.

[6] The use of the Work-Book as a tool of supervision through weekly tests and diagnosis is discussed by F. B. Knight in "Possibilities of Objective Techniques in Supervision," *Journal of Educational Research*, June, 1927.

(3) No provision is made for extensive drill in any particular phase of arithmetic. Provision for extensive diagnosis and drill is an important part of a drill book. A work or drill book is more valuable if it provides for a plan of pointing out specific difficulty, gives help in overcoming it, and affords drill on the right process. This last phase of the test-diagnose-teach-drill system that should be found in a complete work book is not found in this one, especially in reference to individual needs.

The Work, Drill, and Test Sheets [7] are widely used. This material has just undergone a very beneficial revision and should be considered as an important contribution in this field.

An example of a practice test and drill book in algebra is the "Exercises and Tests in Algebra." [8] The table of contents will give some idea of the material presented.

Topic	*Tests*
1. Formulas and Easy Equations	1-36
2. Graphs	37-44
3. Directed Numbers	45-58
4. Elementary Operations (I)	59-84
5. Linear Equations with One Unknown and Problems	85-120
6. Elementary Operations (II)	121-152
7. Fractions	153-174
8. Fractional Equations	175-192
9. Ratio, Proportion, and Variation	193-196
10. Numerical Trigonometry	197-198
11. Simultaneous Linear Equations	199-206
12. Powers and Roots	207-218
13. Quadratic Equations	219-224

This table gives a general classification of the topics covered by the various tests. Review drill, however, is provided all through the book, and certain minor topics are included under two or more general heads.

These exercises and tests in algebra are developed to cover the most important phases of each of the topics listed. Each

[7] Lennes, N. J., *Work, Drill, and Test Sheets*, Laidlaw Bros., 1924.

[8] Smith, Reeve, and Moss, "Exercises and Tests in Algebra." Ginn and Co., 1926.

phase under each topic is tested twice. In the following list of the various tests under the topic "Fractions," it will be noted that each phase of fractions is covered by at least two tests.

Test No.	*What is Tested*
153-154	Reducing Fractions to Lowest Terms
155-156	Fractions and Signs
157-158	Reducing to an Integral or Mixed Expression
159-160	Reducing to a Fraction
161-164	Addition and Subtraction of Fractions
165-166	Multiplication of Fractions
167-168	Division of Fractions
169-170	Complex Fractions
171-172	Principles of Fractions
173-174	Review of Fractions

These practice tests are in general diagnostic. The use of a second practically equivalent test on the same phase of the work gives an excellent opportunity to determine what result has been attained from remedial teaching. Practice tests of this kind should, in general, serve three ends: (1) provide an index of the success of instruction, (2) diagnose class and individual difficulties, (3) provide more effective learning through drill and motivation and self-measurement.

Practice tests in other fields are not so well developed as in mathematics. However, some more or less successful attempts have been made in other subjects. Only a few attempts will be discussed here. A good example of an effort in the field of English is "Practice Leaves in English Fundamentals," Form *A* and Form *B*.[9] Forms *A* and *B* are approximately equivalent forms. Each form contains 30 practice tests covering such phases of English work as No. 1, Sentence Sense; No. 6, Parts of Speech; No. 13, Capitalization; No. 23, Spelling, adding suffixes; No. 27, Punctuation, Series, Responsives, Quotations, Semicolons; No. 30, Review. The use of such material as these Practice Leaves in English should give drill,

[9] Logan, Cleveland, and Hoffman, "Practice Leaves in English Fundamentals for Rapid Drills and Test," D. C. Heath & Co., 1926.

test results, and point out difficulty. These tests are easily checked either by the pupil or by the teacher. A standard progress chart is used with each form.

The following exercise is an example of the idea of practice material as it can be developed for use in geography.[10]

Exercise 5

A list of cities with mean annual temperature.

1.	Charleston, S. C.	65°
2.	New Orleans	69°
3.	New York	52°
4.	Point Barrow	7°
5.	Portland, Me.	45°
6.	Vera Cruz	77°
7.	Winnipeg	35°

A. Cities arranged according to temperature.	B. These cities arranged according to latitude.
1.	1. .....
2.	2. .....
3.	3. .....
4.	4. .....
5.	5. .....
6.	6. .....
7.	7. .....

Above is a list of cities, alphabetically arranged, with their mean annual temperatures. By "mean annual temperature" is meant the average temperature throughout the year. Look at this list carefully. Find the city with the highest temperature. Write the name of this city on line No. 1, under A, where it says, "Cities arranged according to temperature." Find the city with the next highest temperature and write it on line No. 2, and so on until you have all the cities arranged according to temperature, the hottest city first and the coldest city last.

Now look at the map. Find which one of these cities is farthest south. Write the name of this city on line No. 1, under B, where it says, "Cities arranged according to latitude." Find the city next farthest south. Write it on line No. 2, and so on until you have all the cities arranged according to latitude, the city farthest south first and the city farthest north last.

Compare the two lists you have made. What do you find?

Read the following statements and put a + before the one that most nearly represents the most important conclusion that you think should be drawn from the facts presented in this exercise.

[10] Taken from unpublished practice material of H. L. Smith, Velorus Martz, and Wendell W. Wright.

....1. Different cities have different temperatures.
....2. New York is a very large city.
....3. Cities nearer the equator have a higher temperature than those farther away.
....4. The mean annual temperature of Point Barrow is less than that of any other city listed.

In the field of music, a set of notebooks in music appreciation has been developed for use with any regular class text, beginning with the fourth grade. A part of each lesson in appreciation is devoted by the pupil to writing his discoveries in music in his notebook, called "Music Notes."[11] His reactions to the music in which he has participated by singing, etc., or to which he has listened, are thus definitely recorded and can be checked by the teacher for evidence of growth in appreciation of music. If pupils fail to grasp certain phases of the work, opportunity is offered to reteach and recheck.

[11] Books 1, 2, 3, 4 of the series *Music Appreciation for Every Child* and *Music Appreciation for the Junior High School,* by Glenn, Lowry, and De Forest. Published by Silver, Burdett and Company.

CHAPTER XIX

Intelligence Tests

What is Intelligence? The following quotations may give a fair answer to the question from the points of view of the psychologist and of the educator.

"Intelligence is represented in behavior by the capacity of the individual to adjust himself to new situations, to solve new problems, to learn. On the side of descriptive psychology, intelligence is exhibited especially by capacity for carrying on the higher mental operations for abstract thought, for dealing with symbols, for generalizing and for reasoning." [1]

"We may then define intellect in general as the power of good responses from the point of view of truth or fact."—Thorndike.[2]

"An individual is intelligent in proportion as he is able to carry on abstract thinking."—Terman.[2]

"We need only add a useful way of thinking of intelligence, particularly for the teacher, namely, as ability to learn." [3]

"Intelligence more adequately defined would be expressed as the ability of an individual to analyse and synthesize." [4]

These definitions and descriptions of intelligence may differ from those ordinarily given by the layman. The difference is primarily that the layman includes many more factors in his

[1] Frank N. Freeman, "Mental Tests," Houghton Mifflin Co., 1926, p. 491.

[2] Symposium (21) Intelligence and Its Measurement, *Journal of Educational Psychology*, Vol. XII, No. 3: pp. 123-47; March, 1921.

[3] Rudolph Pintner, "Intelligence Testing," Henry Holt & Co., 1925, p. 57.

[4] S. C. Kohs, *Intelligence Measurement*, The Macmillan Company, 1923.

definition than does the psychologist. In addition to ability to learn or to adjust to a new situation, the layman's concept of intelligence includes qualities of leadership, fluency of language, economic success, poise, practicalness, experience, social ability, and education. The psychologist's concept, on the other hand, includes primarily acquiring capacity, mental alertness, and ability to think in a new situation. It is this latter concept of intelligence that the educator tries to measure with intelligence tests. The confusion that has resulted from these varying concepts as to what intelligence is, has brought some adverse criticism of the use of these instruments of measurement called intelligence tests. In view of the fact that only this more limited concept of intelligence is measured by intelligence tests, it would probably have been better had these tests been called by some other name.

Intelligence tests are not an entirely new and isolated development. They are rather an outgrowth of experimental psychology, that has been developing for many years in this and other countries. At first intelligence tests were generally carried on in the psychological laboratories, where they were applied to students in the classes.

The study of ways and means of discovering and selecting backward and feeble-minded children was first worked out in the early part of this century by Binet and Simon. The Binet-Simon test was an individual test, to be given to only one person at a time. The individual test was later followed by group tests of intelligence, which were first used on a large scale during the World War. The use of intelligence tests of both the individual and the group type is now common in our schools. These intelligence tests are measures of the manifestation of capacity in action. They are not a direct measure of capacity. In so far as the product or manifestation of capacity is a true picture of that capacity, the measure is accurate.

The difference between achievement tests and intelligence

tests is that the achievement test attempts to measure the products or results of training, while intelligence tests attempt to measure the original or innate capacity of the pupil to acquire knowledge and to do. Each measure is dependent on the other, and neither is a pure measure. As desirable as it would be to have it so, nevertheless it is not true that intelligence tests are pure measures of capacity uninfluenced by training. The statement that schooling does not influence innate capacity or intelligence may be entirely true. Nevertheless the statement that schooling has no influence on scores made by school pupils on a group intelligence test is false. The influence of school training on an intelligence test score can, no doubt, be kept to a minimum in a good intelligence test. By certain techniques employed by the makers of mental tests, this overlapping of capacity and the influence of training is often kept somewhat low. It is, however, always present to a degree. If, in the development of a test of intelligence, the particular activities called for are either new or based on the common experience of all the persons with whom the test is used, then the influence of training on scores made is kept to a minimum.

That different mental tests do not measure the same thing is shown by the low correlations obtained between the various mental tests given to the same pupils. Each of the following intelligence tests was correlated with every other test given. These thirteen correlations for each test were then reduced to an average correlation. Thus the correlation of .75 for the Terman Group Intelligence Test represents the average of the thirteen individual correlations between this test and the others of the table.

The correlation of intelligence scores with achievement scores has not always been as high as might be expected. This is due to various factors.

(1) Often neither the measure of achievement nor the measure of intelligence is highly reliable. If the true reliability coefficient of an intelligence test is only .90 and the true co-

TABLE 52.—AVERAGE CORRELATIONS OF FOURTEEN GROUP INTELLIGENCE TESTS WITH EACH OF THE OTHER THIRTEEN TESTS [5]

Test	*Average Correlation*
Terman Group Test	.75
National A	.74
Haggerty	.73
Illinois	.72
Otis	.71
Mentimeter	.66
Pressey Survey	.65
National B	.62
Thorndike Reading	.59
Dearborn—1	.58
Pressey Cross-Outs	.56
Dearborn—2	.55
Wylie Opposites	.53
Myers	.46

efficient of reliability of an achievement test is only .90, the highest correlation between the two will probably not be over .81.

(2) The measure of achievement is in many cases not complete. A measure of achievement in arithmetic alone will not give as high a correlation with the measure of intelligence as will the composite measure of achievement in a group of school subjects. The writers have found on careful testing with 150 cases a correlation of $.83 \pm .0178$ between achievement in school work as measured by the Stanford Achievement Test and intelligence as measured by the Otis Self-Administering Intelligence Test. They have also found a correlation of $.81 \pm .024$ between the same intelligence test and the Indiana Composite Achievement Test, which includes nine school subjects. The average correlation between school marks on one subject and intelligence tests scores is, however, usually not much above .60.

(3) Although intelligence tests may accurately measure ca-

[5] G. M. Ruch and George D. Stoddard, *Tests and Measurements in High School Instruction*, World Book Company, 1927. Also R. Franzen, "Attempts at Test Validation," *Journal of Educational Research*, Vol. VI, 1922, pp. 145-58.

pacity to do work, they certainly do not measure the will to do work. The pupil who has capacity to do a high grade of work may be lazy or have little desire to do more than the average of his class. His health may be such as to prevent good work. Naturally his achievement will not in either case be consistent with his mental capacity. The correlation between measures of achievement and measures of intelligence in such instances will consequently be lowered.

(4) Many pupils have developed incorrect habits of work in learning. In many cases these habits have not been corrected. A pupil may have obtained the idea that to add properly he should find all the numbers that, combined, will make ten. Through the application of this method he may become very inefficient in adding. It is fair to assume, therefore, that low achievement is not always due to low intelligence but in many cases is due to bad habits of learning. For this reason pupils do not live up to what their intelligence test scores indicate they should be able to do. Other causes may influence the inconsistency of intelligence and achievement measures, but they are of less importance than the causes named.

THE INDIVIDUAL MENTAL TEST

In an attempt to measure the ability of children, Binet first put together several tests of a type like those used in the laboratory of an experimental psychologist. These he first used with children about 1905. The tests were so arranged that they increased in difficulty. He found that children of a certain age could do some parts of these tests and could not do others, and that certain parts of the complete test represented the type of performance for a certain age. These two ideas—the one of increasing difficulty of material and the other of certain performance or reaction representing a particular age of the child—formed the basis for the development of the mental age scale. Binet's early work has undergone many

revisions. The most important revision was that known as the Stanford Revision.[6]

To give some idea of this individual age scale mental test, a sheet from the 10th year test is given. Directions for giving the 10th year tests are also given.[7]

Directions: The following directions are not complete but will give some idea of the way in which the test is given.

1. *Vocabulary.* Say to the child: "I want to find out how many words you know. Listen; and when I say a word you tell me what it means." The pupil should know 30 definitions.

2. *Absurdities.* Say to the child: "I am going to read a sentence which has something foolish in it, some nonsense. I want you to listen carefully and tell me what is foolish about it."

3. *Designs.* A card with two drawings on it is shown to the pupil. Before showing the card, say: "This card has two drawings on it. I am going to show them to you for ten seconds; then I will take the card away and let you draw from memory what you have seen. Examine both drawings carefully and remember that you have only ten seconds."

4. *Reading and Report.* The copy of selection used by the subject [pupil] is printed in heavy type at the top of the page so that it can be read by the child. (See test on page 446.) The Stanford record booklet contains the selection in a form suitable for use in scoring. (See No. 4 on page 447.) Hand the selection to the subject [pupil], who should be seated comfortably in a good light, and say: "I want you to read this for me as nicely as you can." The subject must read aloud. The subject is not warned in advance that he will be asked to report what he has read, but as soon as he has finished reading, put the selection out of sight and say: "Very well done. Now, I want you to tell me what you read. Begin at the first and tell everything you can remember."

5. *Comprehension.* After getting the pupil's attention, you ask: "What ought you to say when some one asks your opinion

[6] L. M. Terman and H. G. Childs, "A Tentative Revision and Extension of the Binet-Simon Measuring Scale of Intelligence," *Journal of Educational Psychology*, Vol. 3 (1913), pp. 61-74, 133-43, 198-208, 277-89.

[7] L. M. Terman, *The Measurement of Intelligence*, Houghton Mifflin Co., 1916.

about a person you don't know very well?" Ask two additional questions—see "b" and "c" under section 5 of test.

6. *Words.* Say to pupil: "Now I want to see how many different words you can name in three minutes. When I say ready, you must begin and name the words as fast as you can, and I will count them. Do you understand? Be sure to do your very best, and remember that just any words will do."

Following is a sheet[8] of the test booklet used by the examiner. The items marked + indicate that a child did them correctly and received credit for them. The ones marked — were not passed.

New York, September 5th.—A fire last night burned three houses near the center of the city. It took some time to put it out. The loss was fifty thousand dollars, and seventeen families lost their homes. In saving a girl who was asleep in bed a fireman was burned on the hands.

YEAR X

(6 tests, 2 months each, or 4 tests [the ones marked *], 3 months each.)

— * 1. Vocabulary, 30 words. Score. $\frac{25}{}$. Total Vocab.....

— * 2. Absurdities. (4 of 5. Warn. Spontaneous correction allowed.)

+ *a.* "A man said: 'I know a road from my house to the city which is down hill all the way to the city and down hill all the way back home.'"

+ *b.* "An engineer said that the more cars he had on his train the faster he could go."

— *c.* "Yesterday the police found the body of a girl cut into 18 pieces. They believe that she killed herself."

[8] Published by Houghton Mifflin Company, Boston, Mass.

— *d.* "There was a railroad accident yesterday, but it was not very serious. Only 48 people were killed."

e. "A bicycle rider, being thrown from his bicycle in an accident, struck his head against a stone and was instantly killed. They picked him up and carried him to the hospital, and they do not think he will get well again."

\+ 3. *Designs.* (1 correct, 1 half correct. Expose 10 seconds.) *a*..... *b*.....

— 4. Reading and report. (8 memories. 35 seconds and 2 mistakes in reading.)
Memories.... Time for reading.... Mistakes....

New York. | September 5th. | —A fire | last night | burned | three houses | near the center | of the city. | It took some time | to put it out. | The loss | was fifty thousand dollars, | and seventeen families | lost their homes. | In saving | a girl | who was asleep | in bed, | a fireman | was burned | on the hands.

— * 5. Comprehension, 4th degree. (2 of 3. Question may be repeated.)

\+ *a.* "What ought you to say when some one asks your opinion about a person you don't know very well?" .....

— *b.* "What ought you to do before undertaking (beginning) something very important?".....

c. "Why should we judge a person more by his actions than by his words?".....

— * 6. 60 words. (Score half-minutes separately. Illustrate with clouds, dog, chair, happy.) 1.... 2.... 3.... 4.... 5.... 6....
Method .....

A DERIVED SCORE

From tests like the preceding one, the mental age of the individual is figured. The mental age found in this case is

merely a statement of the fact that the particular child tested performed as well on this test as the average child that is ten years and two months old. It is expedient to get a single measure that signifies the relationship between a pupil's performance and what his performance would have been had it been like that of the average pupil for his chronological age. This single term is called the intelligence quotient, or I.Q. This term expressed in another way means the number that shows the relation between the child's mental development and what is expected of a child of his chronological age. This measure is a constant regardless of age[9] and is a means by which children of different ages can be compared. The I.Q. is obtained by dividing the mental age by the chronological age. As a matter of practical convenience, the quotient is multiplied by 100 to clear it of decimal fractions and to express it as a whole number. In the case given, the mental age was 10 years, 2 months; expressed in months, this is 122. The chronological age is 11 years, 2 months; expressed in months, this is 134. Now, by dividing the mental age (M.A.) by the chronological age (C.A.) and then multiplying by 100, the result is:

$$\frac{\text{M.A.}}{\text{C.A.}} = Q \times 100 = \text{I.Q.}$$

$$\frac{122}{134} = .91 \times 100 = 91$$

The I.Q. of this pupil is 91.

A study of Table 53 will give some idea of the significance of an I.Q. of 91.

TABLE 53.—THE DISTRIBUTION OF INTELLIGENCE QUOTIENTS[10]

The percentage of individuals who have various intelligence quotients or lower:

I.Q.	70	73	76	78	85	88	91	92	95
Per cent...	1	2	3	5	10	15	20	25	33.3

[9] This applies to pupils with a chronological age of fourteen or less.
[10] Frank N. Freeman, *Mental Tests*, Houghton Mifflin Co., pp. 99-100.

The percentage of individuals who have various intelligence quotients or higher:

I.Q.	106	108	110	113	116	122	125	128	130
Per cent...	33.3	25	20	15	10	5	3	2	1

This table is to be read as follows: The lowest one per cent of persons in general have an I.Q. of 70 or below. The lowest two per cent have an I.Q. of 73 or below. The lowest twenty-five have an I.Q. of 92 or below. The highest one per cent have an I.Q. of 130 or above. The highest twenty per cent have an I.Q. of 110 or above, and the highest thirty-three and one-third per cent have an I.Q. of 106 or above. We see that according to this table the middle fifty per cent of all individuals have intelligence quotients ranging from 92 to 108.[11]

Another way of determining the significance of a particular I.Q. would be to compare it to the following descriptive table, Table 54. It will be noted that more than 50 per cent of the people are in the range of I.Q. marked "normal" in the chart, Table 53, while not more than 1 per cent of individuals fall in the last three classes of Moron, Imbecile, and Idiot.

TABLE 54.—THE SIGNIFICANCE OF THE INTELLIGENCE QUOTIENT [11]

Class	*Range of I.Q.'s*
Near genius	140+
Very superior	120-140
Superior	110-120
Normal	90-110
Dull	80-90
Border-line	70-80
Moron	50-70
Imbecile	25-50
Idiot	0-25

The individual test, Stanford Revision of the Simon-Binet, if properly given, is an accurate measure of intelligence of children who speak English. It is a valuable mental measure. It is the standard by which other tests are usually evaluated. Because the test has often been given by those who are not proficient in its use, the value of its results has sometimes been questioned. It is doubtful whether results obtained from giv-

[11] *Ibid.*

ing the test should be taken as an accurate measure unless the examiner is well trained in its use. The test, if carefully given, is especially valuable as a clinical instrument when used with many other scientific data and with observation of the individual. It is also of great value in selecting cases for special classes in the public schools. Recruits for classes for inferior and for superior pupils should be selected at least partially by the use of a measure of intelligence. Since the Stanford-Binet test can be given to only one individual at a time, since it requires from 40 to 60 minutes to give and score, and should furthermore be given by an expert, the use of this scale of measurement may well be reserved for special cases.

Because of the fact that performance in the Stanford Revision of the Binet-Simon test is dependent to some degree upon language ability as well as upon general intelligence, it is best at times to supplement the Binet test with another type of intelligence test that is not so dependent upon verbal response. Such a test as the Pintner-Patterson Performance Scale [12] has been developed and used in the testing of foreign and deaf pupils as well as English-speaking pupils. This scale consists of 15 separate tests, in which the pupil is to do something rather than to say something. For example, the test known as the Ship Test consists of the picture of a ship cut into ten pieces of the same size and shape, which are to be fitted together properly to make a ship in a rectangular frame. The test is scored on the quality of performance. Other tests that deal primarily with analysis of mental capacity and that are used mostly outside the schoolroom are the Healy-Fernald Tests, used with defective children.

The tests just described are used in the main as individual mental tests.

[12] R. Pintner and O. G. Patterson, *A Scale of Performance Tests*, D. Appleton & Co., 1917.

GROUP INTELLIGENCE TESTS

The most widely used intelligence tests are those that may be given to all of a group of pupils at the same time. These group intelligence tests, like the individual intelligence tests, have been developed from tests in use many years in the experimental psychology laboratories. The Army Group Intelligence Scale Alpha was one of the first group tests widely used. It was used in the army for the purpose of quickly selecting and classifying recruits. This test is suitable to use with men who understand and read English. Another group test called Beta was given to those men who did not understand English or who were illiterate. It consisted of pictures and diagrams, with which the men worked after receiving the directions in pantomime. If a very low score was made by a soldier on the group tests, he was given the Stanford-Binet test or a performance test like the Pintner-Patterson Performance Scale described on page 450.

To give some idea of the tests, parts of the Terman Group Test of Mental Ability,[13] for grades 7 to 12, Form B, are given here.

EXAMINATION: FORM B

Test 1. Information

Draw a line under the *one* word that makes the sentence true, as shown in the sample.

Sample: Our first President was
Adams Jefferson Lincoln Washington

1. The most gold is produced in
Alaska Tennessee Texas New York 1
2. A peck is a fourth of a
barrel bushel gallon keg 2

[13] Prepared by L. M. Terman and published by the World Book Company, 1920.

3. The Yale is a kind of
   screw lock hammer wrench 3
4. Chalk is a kind of
   flour limestone slate marble 4
5. Among birds that migrate are
   eagles hawks owls robins 5

Test 2. Best Answer

Read each question or statement and make a cross before the *best* answer, as shown in the sample.

Sample:
Why do we buy clocks? Because
1. We like to hear them strike.
2. They have hands.
× 3. They tell us the time.

1. We should "think twice before we speak," because
   1. We may think of more things to say.
   2. We are then more sure to say the right thing.
   3. If we speak too quickly, we may stammer.
2. The saying, "Idle brains are the devil's workhouse," means
   1. The devil works with his brains.
   2. People should not work for the devil.
   3. People who are idle get into trouble.

Test 3. Word Meaning

When two words mean the *same,* draw a line under "same." When they mean the *opposite,* draw a line under "opposite."

Samples:
fall—drop ..... same—opposite
north—south ..... same—opposite

1. alert—sluggish ..... same—opposite 1
2. active—passive ..... same—opposite 2
3. procure—obtain ..... same—opposite 3
4. minimum—maximum ..... same—opposite 4
5. kindle—quench ..... same—opposite 5
6. hazardous—dangerous ..... same—opposite 6

Test 4. Logical Selection

In each sentence draw a line under the *two* words that tell what the thing *always* has. Underline *two,* and *only two,* in each line.

Sample: A man always has
body cap gloves mouth money

1. A snake always has
poison rattle stripes tail tongue 1
2. A bicycle always has
brakes frame rubber pump wheels 2
3. A box always has
depth hinge lid sides wood 3
4. Food always has
nutriment salt starch sweetness taste 4
5. A soldier always has
bayonet commander duty flag tent 5

Test 5. Arithmetic

Find the answers as quickly as you can. Write the answers on the dotted lines. Use the bottom of the page to figure on.

1. Frank has 12 marbles. He bought 3 more, and then lost 6. How many had he left? *Answer*.....
2. What number multiplied by 16 equals 24×2? *Answer*.....
3. A man bought some sheep for $150. He sold them for $200, gaining $5 per head. How many did he buy? *Answer*.....

Test 6. Sentence Meaning

Draw a line under the right answer, as shown in the samples.

Samples: { Is coal obtained from mines?.. Yes No
Are all men six feet tall?..... Yes No

1. Are cartoons made by cameras?..... Yes No 1
2. Are transparent substances used in windows?..... Yes No 2

3. Do hoboes ever wear dilapidated garments? Yes No 3
4. Is burlap a kind of lumber?..... Yes No 4
5. Do hermits usually live in seclusion?..... Yes No 5

Test 7. Analogies

Samples:
Ear is to hear as eye is to
table see hand play
Hat is to head as shoe is to
arm coat foot leg

Do them all like the samples.

1. Picture is to see as sound is to
noise music hear bark 1
2. Uncle is to nephew as aunt is to
brother sister niece cousin 2
3. Add is to subtract as multiply is to
add divide arithmetic increase 3
4. Shell is to nut as skin is to
person soft white coarse 4
5. Tree is to forest as person is to
couple men women crowd 5

Test 8. Mixed Sentences

The words in each sentence below are mixed up. If what a sentence means is *true,* draw a line under "true." If what it means is *false,* draw a line under "false."

Samples:
hear are with to ears..... true false
eat gunpowder to good is... true false

1. countries several produced wheat in is... true false 1
2. pays cautious it be to often..... true false 2
3. north all railroads south and run..... true false 3
4. men industrious pay good should get.... true false 4

Test 9. Classification

Samples:
1. bullet cannon gun sword pencil
2. Canada Chicago China India France

In each line cross out the word that does not belong there. Cross out *just one word* in each line.

1.	elm	brier	maple	oak	poplar	1
2.	needle	pan	stitch	thimble	thread	2
3.	Governor	King	Mayor	President	Priest	3
4.	baby	calf	colt	doll	kitten	4

Test 10. Number Series

Samples:

5	10	15	20	25	30 ..	35 ..
20	18	16	14	12	10 ..	8 ..

In each row try to find out how the numbers are made up, then on the two dotted lines write the *two* numbers that should come next.

1st Row	3	4	5	6	7	8	....	....
2d Row	3	6	9	12	15	18	....	....
3d Row	10.8	9.7	8.6	7.5	6.4	5.3	....	....

This test is highly reliable. It is easy to give and to score. The giving of the test requires approximately 30 to 35 minutes.

A test of intelligence applicable to lower grades is the Kingsbury Primary Group Intelligence Scale,[14] Form *A*, from which selections from the instructions are given.

INSTRUCTIONS

I. Right Answers Test

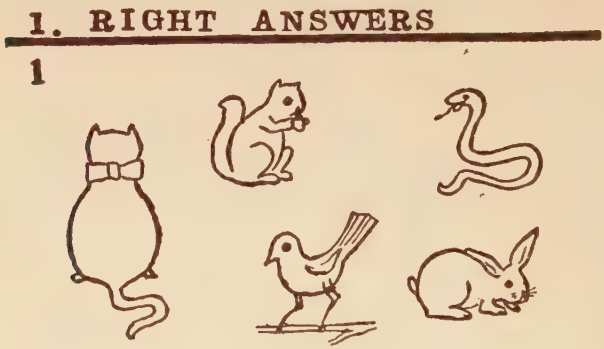
"As I walked through the woods one day, I saw these five animals. (See page 456.) One was *flying* from tree to tree. Draw a circle around the picture of the one that was flying from tree to tree.

"Which should it be? Why the bird, of course, because that is the only one that can fly. Take your pencils and mark a circle around the bird, like this (mark). It doesn't have to be a good circle,—just a ring around it, so I can tell that you

[14] Published by the Public School Publishing Company.

know which one is right. Now don't mark any more till I tell you what to do. As I tell you the rest, you will have to work as quickly as you can without making a mistake. Don't answer aloud but just answer by marking in the right place."

Example I



II. Opposites Test

"Here are some queer pictures, several rows of them, with one picture in the square at the beginning of each row. In each row there is one picture which is more different from the one in the square than any of the others. That is the same as saying it is the opposite of the one in the square. . . . Now look at the second row. See, in this square we have a black spot; then after it come three other spots. Which of these other three spots is most different from the black spot? Now draw a circle around it."

Example II

III. Completion Test

"Here are several rows of drawings of different kinds. In each row there is a blank space—sometimes two blank spaces—

where a drawing has been left out, which is needed to finish the row. Make up your mind what is left out from the blank space, and draw in whatever is left out so that the row will be finished."

Example III

On page 458 is a list of group tests that are highly valuable for school use, with helpful information concerning each.

These group scales are of course only approximate measuring instruments. I.Q.'s derived from them are less likely to be stable than with such a test as the Stanford-Binet test if it is carefully given. The value of these group tests lies largely in the fact that they give a good approximate measure of ability in rather a short time. It is fair to assume that they are not pure measures of intelligence and that the results obtained from their use should therefore not be taken alone as a criterion of capacity, although they are good indicators of it.

There are four important uses which the school teacher, supervisor, or administrator may make of intelligence tests.

1. Many more children fail to be promoted from the first grade than from any of the grades of the school. This is no doubt due to many factors such as health, previous home training, attendance, etc. Dickson reports [15] that of 261 children tested in the lower grades only 3 with an I.Q. above 110 failed to be promoted; whereas, of 129 who failed of promotion at least once, 84 had an I.Q. below 90. Bright younger children are capable of doing first grade work better than the average six-year-old. It might be advisable to allow children to begin first-grade work when they have developed a mental age of

[15] "The Use of Mental Tests in School Administration," Board of Education, Berkeley, California, 1922.

TABLE 55.—GROUP INTELLIGENCE TESTS SUGGESTED FOR USE

Name of Test	*Author*	*Grades*	*Manual*	*Forms*	*Time to Give*	*Norms*	*Reliability*
Pintner-Cunningham Primary Mental Test [a]	Rudolph Pintner Bess V. Cunningham	K, 1, 2, 3	Yes	1	30–45 min.	Age & grade	r. .89 r. .93
Dearborn Group Series 1 [b]	Walter F. Dearborn	1–3	Yes	1	45 min.	Age	
Detroit First Grade [a]	Anna M. Engel	1	Yes	1	20–25 min.	Grade	
Kingsbury Primary [c]	Forrest A. Kingsbury	1	Yes	1		Age	
National Intelligence Scale A and B [a]	R. M. Yerkes and others	3–8	Yes	2	30–55 min.	Age, grade; (sex in preparation)	
Otis Intermediate Self-Administering [a]	Arthur S. Otis	4–9	Yes	2	35 min.	Age, grade	.948
Army Alpha [d]	Div. of Psychology Medical Dept. U. S. A.	9–adult	Yes	5	40–50 min.	Adult	.95
Terman Group Test of Mental Ability [a]	Lewis M. Terman	7–12	Yes	2		M.A. grade	About .90
Thorndike Intelligence Examination for High School Graduates [e]	E. L. Thorndike	H. S. Seniors and College Freshmen	Yes	Indef.	3 hours	Rough sta. for College Entrance	.85

a. World Book Co.
b. J. B. Lippincott Co.
c. Public School Publishing Co.
d. Division of Psychology, Medical Department, U. S. A.
e. Bureau of Publications, Teachers College, Columbia University

5 years and 6 months to 6 years. Certainly some distinction should be made upon the basis of the ability to do the beginning work of the elementary school. This classification of children as well as differentiation of the course of study to meet the demands of individual differences can best be made, with children beginning school work, through the use of intelligence tests.

2. In selecting children for special classes, intelligence tests can be of very great help. Whether such classes are to be for backward children or for gifted children, the selection of pupils for them can be more accurately made through the use of intelligence tests supplemented by other available information than by any other method. In the selection of these children for special classes it is advisable to use an individual examination also.

3. Intelligence tests are widely used in classifying pupils into ability groups for instructional purposes. The general practice is to group pupils into three groups on the basis of ability. One of these groups contains the bright pupils, another the pupils of average ability, and the third the slow pupils. As such a grouping is somewhat rough and arbitrary, opportunity to change pupils from one group to another should be given from time to time. The advantage of this kind of grouping, which may be called homogeneous grouping, is that it offers an opportunity for different rates of progress over the required work. It gives opportunity for adjustment and enrichment of the course of study according to the group. Methods of teaching, too, may be better adjusted to the type of pupils. The advantages of homogeneous grouping have not always been realized. We are badly in need of more scientific knowledge concerning the content of the course to be offered to the various groups and concerning the methods to be used in teaching the various groups. Classification of pupils into homogeneous groups does not produce more effective learning on the part

of the pupils unless the problems of what they should be taught and how they should be taught are attacked at the same time.

Mental tests should not serve as the only basis for the classification of pupils. Teachers' judgments, industry, the health of the pupil, previous school marks, and other factors should also be considered in making any decision regarding classification.

4. In educational and vocational guidance, especially in the junior high school, mental tests can be of much aid. They will serve as guides in the selection of further courses and in the educational plan that will lead to certain vocations. The results of a mental test of a pupil will, however, need to be supplemented with much other information. The following two examples will give some idea of the manner in which this guidance may be given.[16]

Roe, Richard.

Score Army Scale—150.
Army Scale mental age: 17 years, 4 mos.
Army Scale I.Q.—120.

High school subjects which pupil desires to take:
English
History
Algebra
French

Chronological age: 14 years, 4 mos.
Stanford-Binet mental age: 16 years, 9 mos.

Stanford-Binet I.Q.—117.

Educational plans: To finish high school then attend a university or the U. S. Naval Academy. Vocational ambition: Chemical engineer or naval officer.

Grade of work done in elementary and intermediate schools: Very poor. Estimated as "average" by some grade teachers, and as "below average" by others. Comment of examiner: "Boy has ability but needs to be waked up. Suggest that he take general science in place of history for first year. Also suggest that he be placed in first division in algebra where he will have to work. He will need to develop ability in both science and mathematics if he is to follow his vocational ambition."

[16] W. M. Proctor, "Psychological Tests and Educational Guidance," *Journal of Educational Research*, Vol. 1 (1920), pp. 369-81.

Brown, Carrie.

Score Army Scale—100.
Army Scale mental age: 14 years, 0 mos.
Army Scale I.Q.—89.

Chronological age: 15 years, 7 mos.
Stanford-Binet mental age: 14 years, 2 mos.
Stanford-Binet I.Q.—90.

High school subjects which pupil desires to take:
English
Algebra
Latin
Typing
Drawing

Educational plans: Go to Mills College.
Vocational ambition: To be a chemist.

Grade of work done in intermediate and grammar schools: Grades in 8A class only fair, even in work that is being repeated. Estimates of elementary and intermediate teachers: "slow" but a conscientious worker. Comment of examiner: "Should be discouraged from taking Latin. Algebra doubtful, but if she insists in view of desire to go to college assign to second division."

Purposely little space in this book has been given to intelligence tests. However, since intelligence tests are basic to the interpretation of achievement it has seemed advisable in this chapter to make some reference to their development and use.

BIBLIOGRAPHY

Bronner, A. F. and others. *A Manual of Individual Mental Tests and Testing,* Little, Brown & Co., 1927.

Cunningham, Kenneth S. *The Measurement of Early Levels of Intelligence.* Contributions to Education, No. 259, Teachers College, Columbia University, 1927.

Dickson, Virgil E. *Mental Tests and the Classroom Teacher,* World Book Company, 1923.

Dougherty, Mary L. *A Comparative Study of Nine Group Tests of Intelligence of Primary Grades.* Johns Hopkins Press, 1928.

Freeman, Frank N. *Mental Tests,* Houghton Mifflin Co., 1926.

Hines, Harlan C. *Measuring Intelligence,* Houghton Mifflin Company, 1923.

Hollingworth, H. L. *Vocational Psychology,* D. Appleton & Company, 1916.

Kohs, S. C. *Intelligence Measurement,* The Macmillan Company, 1927.

McPhail, Andrew H. *The Intelligence of College Students,* Warwick & York, 1924.

National Society for the Study of Education. *Twenty-first Yearbook, Intelligence Tests and Their Use,* Public School Publishing Company, 1922.

Peterson, J. C. *Early Conceptions and Tests of Intelligence,* World Book Company, 1925.

Pintner, R. *Intelligence Testing,* Henry Holt & Company, 1923.

Ryan, Heber Hinds, and Crecelius, Philipine. *Ability Grouping in the Junior High School,* Harcourt, Brace, 1927.

Terman, L. M. *The Measurement of Intelligence,* Houghton Mifflin Company, 1916.

Terman, L. M. *The Intelligence of School Children,* Houghton Mifflin Company, 1919.

Thorndike, E. L. and others. *The Measurement of Intelligence,* Teachers College, Columbia University, 1927.

Wells, F. L. *Mental Tests in Clinical Practice,* World Book Company, 1927.

Whipple, G. M. *Manual of Mental and Physical Tests,* Second edition, Warwick & York, 1915.

Wood, Ben D. *Measurement in Higher Education,* World Book Company, 1923.

Woodrow, Herbert. *Brightness and Dullness in Children,* J. B. Lippincott Company, 1919.

Yerkes, R. M., editor. *Psychological Examining in the United States Army,* National Academy of Sciences, Vol. XV, 1921.

Yerkes, R. M. and Foster, Josephine C. *A Point Scale for Measuring Mental Ability,* 1923 Revision, Warwick & York.

Yoakum, C. S. and Yerkes, R. M. *Army Mental Tests,* Henry Holt & Company, 1920.

CHAPTER XX

Prognosis and Special Ability Tests

The purpose of a prognostic or a special aptitude test is to predict the possibility or the probability of future success in any particular subject or subjects or in specific types of work. Such a test is an attempt to measure scientifically various factors at one time, and from the results of the measure of these factors to foretell the success of a pupil in future school work or in work outside of school. Not much has been attempted along the line of forecasting success outside of school. Prognostic tests are based on the following assumptions.

(1) Success in a given line of work is dependent upon certain knowledge that is supposed to have been acquired in previous training. Success in Latin, for example, may be dependent upon knowledge of English Grammar. Such grammar has supposedly been taught previous to the beginning of the work in Latin.

(2) Success in a given line of work is dependent upon general intelligence or ability. From the results of a measure of this general ability a prediction may be made.

(3) Success in a given line of work is dependent upon special aptitudes or abilities. For instance, the special abilities necessary to the learning of music may be present to a greater degree in some people than in others. The results of the measure of this special ability or aptitude may be used for predicting success in a given line of endeavor.

(4) Success in a given line of work is dependent upon various traits of personality or character, upon interest, upon certain habits of work, or upon physical health. Other things being equal, the pupil who has stick-to-it-iveness, who works

hard, who is interested, and who has the vitality of a healthy body can be supposed to do more than the pupil who does not have these traits. A measure or estimate of the degree of presence or absence of these traits may serve as a basis of prediction, especially when used with other measures.

(5) Success in a given line of work is somewhat dependent upon the environment in which the pupil lives and works. Other things being equal, if the parents are educated and are interested in the child's school work, and if the conditions for study are favorable and the length of time for study sufficient, the pupil will be more likely to succeed in school work than if these factors are not present. Absence from school might easily influence success in school work. Almost any teacher has observed how a serious disturbance in the home may cause the child's efficiency in school work to be lowered.

The measurement of the various factors that should be taken into account in attempting to forecast or predict future success is a problem that has by no means been solved. Certain questions need yet to be answered in relation to each of the foregoing assumptions.

(1) If success in a given line of work is somewhat dependent upon knowledge previously learned, then the question is, "What is that knowledge that is most important for successful future work?" The statement is often made that success in beginning Latin (9th grade in an 8-4 system) is somewhat dependent upon knowledge of English grammar. Are all of the various phases of English grammar equally important in the learning of Latin? If the knowledge of all parts of English grammar are not equally important, what phases are most important? What mathematics is needed to do the work of high school physics? What knowledge of mathematics is necessary to the proper training of machine shop workers? When such information is determined for any subject, then probable future success in that subject can be measured. A test of this type would be an achievement test of that knowledge considered

basic in the subject in which the pupil is beginning his work. Of course, it is understood that should pupils not have this knowledge the teachers of the Latin or Physics or Machine Shop classes could take time to teach it. However, the assumption here is that the teachers of these classes will go ahead, presuming that such work has been previously taught and learned.

(2) If success in a line of work is somewhat dependent upon general ability, what is the best way to test general ability? Intelligence tests are the best present measure of this particular factor. However, not all intelligence tests measure the same phases of general intelligence. Neither are all such tests equally reliable. The use of more than one good intelligence test is often suggested as a means of obtaining a better measure of general ability. Teachers' estimates of pupils' ability are frequently used to supplement the tests. Does general ability as measured by intelligence tests or teacher's judgment influence success in all subjects? Within certain limits it is likely that this general ability so measured is not a factor in achievement in handwriting.

(3) Much work has been done in attempting to develop measures of special abilities and aptitudes on the assumption that success in any given line of work is dependent upon these abilities and aptitudes. How do these special abilities differ from the unitary abilities that make up what is called general ability or intelligence? How much are the abilities specialized? If it is found that a general group intelligence test will forecast achievement in algebra about as well as a special test of mathematical ability, then the question of the importance of special abilities would need to be answered before special ability tests could be of any real use. Stenquist and Toops agree that the special ability known as mechanical ability has little relationship to general ability as measured by intelligence tests. To what extent are such special abilities innate and unchangeable and to what extent are they merely matters of in-

terest and experience? A child's ability to distinguish between the intensity of two sounds is probably somewhat fixed and innate. In so far as such a power is a necessary part of the pupil's ability to learn music, it can be said that the measure is not influenced by interest or experience. On the other hand, ability to assemble a door lock as part of the test of mechanical ability of a child might be greatly influenced by his previous experience with door locks.

(4) That success in a given line of work is dependent upon interest, traits of personality or character, habits of work, and physical vitality, is readily admitted. How can these traits be adequately measured? This is a most difficult question. At present the measures are very rough. They are usually obtained through ratings made by teachers. The value of these ratings as true measures depends somewhat upon the teacher's acquaintance with the pupil and upon her knowledge of how to use a rating scale. How can physical vitality be measured? At present only the most important physical defects are noted in examinations. The height-weight index is used to give a rough measure of physical efficiency.

(5) Little is scientifically known of the effect of environmental conditions on success in a given line of work.

Two big questions concerning prognosis yet remain. (1) What is success in a given line of work? (2) Of what relative importance are the various factors in forecasting success in a given line of work?

Suppose that adequate measures were available to measure each of the factors mentioned. Suppose the various measures gave the following results for ten students beginning Latin (Table 56).

The table should be read as follows. Pupil A received a rating of 1 (highest) in the measure of necessary previous achievement. Pupil B received a rating of 2, etc. In the theoretical results given, pupil A received the highest rating in the various factors. If he received the same teaching in

Latin as each of the other nine pupils, he should achieve the best of any of the ten pupils in Latin. Pupil J should under the same teaching rank lowest in his achievement. Each of the others should maintain his relative rank in achievement in Latin. It is probable that this would really work out if the various pupils represented the ordinary range of measures in these traits.

TABLE 56.—THEORETICAL RATING OF PUPILS IN VARIOUS MEASURES OF VALUE IN PROGNOSIS

Pupil	*Ratings Obtained from True Measures in*						
	Previous necessary achievement	*General ability*	*Special ability*	*Traits of personality*	*Health*	*Environment*	*Success in H. S. Latin*
A.....	1	1	1	1	1	1	= 1
B.....	2	2	2	2	2	2	= 2
C.....	3	3	3	3	3	3	= 3
D.....	4	4	4	4	4	4	= 4
E.....	5	5	5	5	5	5	= 5
F.....	6	6	6	6	6	6	= 6
G.....	7	7	7	7	7	7	= 7
H.....	8	8	8	8	8	8	= 8
I.....	9	9	9	9	9	9	= 9
J.....	10	10	10	10	10	10	= 10

Now suppose that, instead of the results given in Table 56, the results shown in Table 57 were obtained.

In this table it will be noted that pupil A has a rating of 1 (highest) in two of the measures and a rating of 10 (lowest) in another. Which one is most important in determining what he will do in his Latin work? Will pupil G, whose ratings on the average are as good and whose rating is 1 (highest) in "Environment" and "Traits of personality," do as well as pupil A in his Latin? This question of the relative value of the various traits can be answered only by experimentation, that is, by finding which traits correlate or seem to go with success in Latin work. By further statistical treatment the relative value or weight of each measure can be determined.

The value of any prognostic measure is determined by how nearly the pupil's later success is like the previous forecast or prediction of his success. What determines success in a school subject or in a job? Do teachers' marks or scores on standardized tests indicate success? While success in algebra might be determined by the amount of growth made in traits of

TABLE 57.—THEORETICAL RATING OF PUPILS IN VARIOUS MEASURES OF VALUE IN PROGNOSIS

Pupil	*Ratings Obtained from True Measures in*						
	Previous necessary achievement	*General ability*	*Special ability*	*Traits of personality*	*Health*	*Environment*	*Success in H. S. Latin*
A.....	1	1	4	10	9	3	= ?
B.....	4	2	3	8	7	6	= ?
C.....	6	3	6	7	4	5	= ?
D.....	5	4	2	9	1	10	= ?
E.....	8	5	1	4	8	2	= ?
F.....	10	6	9	2	10	8	= ?
G.....	9	7	8	1	2	1	= ?
H.....	2	8	10	3	3	4	= ?
I.....	3	9	5	5	5	9	= ?
J.....	7	10	7	6	6	7	= ?

character or in physical fitness or in habits of work, the type and amount of algebra learned constitute more nearly the measure of success as it is thought of in the classroom. Somers,[1] in predicting the success of prospective teachers, used as one measure of success or final criterion of success a rating of the teacher on the job. Using a carefully worked out rating scale, Somers obtained judgments of the teacher's success in her first year of teaching from her principal, superintendent, and supervisor. This procedure illustrates another typical method of determining success.

Following is a partial list of Prognostic and Aptitude Tests.

[1] "Pedagogical Prognosis," Teachers College, Columbia University, Contributions to Education, No. 140.

Subjects or Abilities Tested	*Name of Test*	*Range*
English	Aly Placement Test in English for College Freshmen [2]	College Freshmen
	Iowa Placement Examination in English [3]	College Freshmen
Mathematics	Iowa Placement Examination in Mathematics [3]	College Freshmen
	Rogers' Test for Mathematical Ability [4]	Grade 9
Foreign Language	Iowa Placement Examination in Foreign Language [3]	College Freshmen
	Wilkins' Prognosis Test in Modern Foreign Language [5]	H. S. and College
	Morris-Ward Prognosis Test in Latin [6]	Grades 7-10
	Orleans-Solomon Latin Prognosis Test [5]	Beginning H. S. Latin
	Wyckoff Latin Prognosis Test [7]	Beginning H. S. Latin
	Iowa Placement Examination in French [3]	College Freshmen
	Iowa Placement Examination in Spanish [3]	College Freshmen
Science	Iowa Placement Examination in Chemistry [3]	College Freshmen
	Iowa Placement Examination in Physics [3]	College Freshmen
Commercial	Business Aptitude Test [8]	Junior and Senior High School
	Ruggles Test of Clerical Aptitude [9]	High School
	Hoke Prognostic Test of Stenographic Ability [10]	Pupils studying shorthand
Vocational	Kelley Constructive Ability Test [11]	No information
	MacQuarrie Test for Mechanical Ability [12]	Age 10 or older

[2] Bower Aly, Teachers College, Cape Girardeau, Missouri.
[3] Extension Division, University of Iowa.
[4] Bureau of Publications, Teachers College, Columbia University.
[5] World Book Co.
[6] J. V. L. Morris, Teachers College, Alva, Okla.
[7] R. T. Wyckoff, Berea College, Berea, Ky.
[8] Center for Psychological Service.
[9] Bureau of Educational Research, University of Oklahoma.
[10] Gregg Publishing Co.
[11] C. H. Stoelting Co.
[12] Research Service Co.

Subjects or Abilities Tested	*Name of Test*	*Range*
	O'Rourke Series of Aids in Placement and Guidance [13]	Grade 7 through college, and adult
	Stenquist Assembling Tests of General Mechanical Ability [11]	Grade 3 through high school, and adult
	Stenquist Mechanical Aptitude Tests [5]	Grade 6 through college
Professional	Teaching Aptitude Test [8]	Teachers and prospective teachers
	Law Aptitude Examination [14]	Law students

To give some idea of a prognosis test, the Orleans-Solomon Latin Prognosis Test [15] will here be discussed in detail. The authors of this test believe that five factors should be considered in forecasting pupils' success in the study of Latin. These factors are:

(1) His ability to handle Latin situations such as he must meet in his study of Latin; (2) his knowledge of grammar and such other information as is necessary for learning Latin by the methods now in common use, and his skill in applying such information; (3) his school habits: viz., regularity of home work, attentiveness in the classroom, originality, initiative, perseverance, etc.; (4) the environmental factors that influence his school work: viz., number of children in the family, home facilities for study, number of hours a day spent in remunerative work, home incentive to school success, etc.; and possibly (5) certain phases of general intelligence not included under Items 1 and 3 above, that might be measured by a test of general intelligence.[16]

This test attempts to measure the first two factors named. Most of the test is devoted to a measure of "ability to meet Latin situations." To measure this ability, a lesson is assigned, and the pupil is given a certain amount of time to study the lesson; then he is given a test over that lesson. This is a repre-

[13] Educational and Personnel Publishing Co..
[14] West Publishing Co.
[15] Published by the World Book Co., 1926.
[16] From the Manual of Directions, Orleans-Solomon Latin Prognosis Test.

sentative miniature Latin learning situation of first-year Latin. A lesson and the test that follows it are reproduced.

LESSON

I. Examine these words:

puella—the girl
regina—the queen
femina—the woman
Cornelia—Cornelia
amat—likes
culpat—blames

II. Carefully study these sentences with their translations:

1.	**Regina puell*am*** **amat.**	The queen likes the girl.
2.	**Puella regin*am*** amat.	The girl likes the queen.
3.	**Femina Cornelia*m*** **culpat.**	The woman blames Cornelia.
4.	**Cornelia femin*am*** **culpat**	Cornelia blames the woman.

III. Draw one line under the subject of each Latin sentence. Draw two lines under the direct object of each Latin sentence.

Notice the difference between the endings of the subject and the object. Can you tell what the ending of the word will be if you know how it is used in the sentence? Make sure that you know this or you will not be able to do the next test.

Think what this lesson teaches you.

Test 4

Think of what you learned on the preceding page.

These words are to be used in translating the following sentences:

the water—**aqua**
the eagle—**aquila**
the mistress—**domina**
the woman—**femina**
the hen—**gallina**
the maidservant—**serva**
the forest—**silva**
the land—**terra**
likes—**amat**
destroys—**delet**
praises—**laudat**
kills—**necat**
carries—**portat**

Translate these sentences into Latin. Study the lesson again if you need to.

1. The woman likes the forest.....
2. Cornelia likes the woman.....
3. The eagle destroys the land.....
4. The eagle kills the hen.....
5. The mistress praises the maidservant.....
6. The maidservant carries the water.....

Number right ( ) × 2 = ... (*Score*)

The authors of the test say that the study of the first half-year of Latin involves:

(1) The appreciation of inflection corresponding to changes in gender, number, person, tense, and the like; (2) the ability to learn the inflections and to apply this knowledge; (3) the appreciation of differences in usage and ability to employ the appropriate forms; (4) the ability to relate Latin words with derived English words and vice versa; (5) the ability to memorize and recall Latin words and their meanings. Each of these is represented by one or more parts in the prognosis test.[17]

The test can be given before the beginning of class work in Latin. Classification of pupils on the basis of the results can be carried out as a material aid in adjusting the instruction to the needs of the pupil. The test is of course not a perfect measuring instrument and must be supplemented by other measures. Intelligence test scores, grammar test scores, and perhaps more than any other single item school habits supplement the results of the prognosis test. The authors of this test have worked out a chart, which is reproduced in Table 58.

The results obtained from the use of this chart are worth much when added to the rating obtained from the test.

The total time that it takes to give the test is approximately one hour. The Manual of Directions is good.

Another method of developing a prognosis test was used by

[17] From Manual of Directions, Orleans-Solomon Latin Prognosis Test.

TABLE 58.—SCHOOL-HABIT RATING SCALE[18]

Below is a list of characteristics which make for success in school work.

Rate the pupil on each of these traits. After each trait is a line, and below the line the trait is described at the three points—the poorest, average, and best of each trait. On each line make a cross to indicate your estimate of the extent to which this pupil possesses each trait. You do not have to make a cross on one of the three points indicated but may make it anywhere along the line.

ATTENTION		
Extreme inability to put attention on task.	Usually pays attention; can be distracted.	Difficult to distract him while studying or during recitation.
NEATNESS		
Exceedingly careless in written work.	Written work fair in general appearance.	Unusually painstaking in general appearance and in details.
HONESTY		
Cheats at every opportunity.	Might cheat if tempted.	Could not be tempted to cheat.
INTEREST		
Shows no interest in any school work.	Can be interested in school work by use of some incentive.	Genuinely interested in all school work for its own sake.
INDUSTRY		
Exceedingly lazy. Lacks energy to work even with most interesting subject matter.	Will work with ordinary incentive. No unusual exertion except on special occasions.	Incessant worker. Needs no urging.
INITIATIVE		
Takes no leading part in any class activity even when urged.	Occasionally tries to lead in class activities and is pleased when he does, but is usually satisfied to follow.	Always trying to lead in all class activities.
AMBITION		
Does not care at all whether he gets passing mark, or whether he learns anything.	Satisfied with average mark and promotion. Pleased with good marks but will not exert himself for them.	Exceedingly anxious to get highest possible standing in all school work.

[18] *Ibid.*

PERSISTENCE	Gives up working as soon as he meets the slightest difficulty.	Will keep on working in face of difficulty but not very long.	Will not give up until he has solved all difficulties or is exhausted.
RESPONSIBILITY	Never does what he promises; never does assigned work.	Fairly regular in doing home work. Will be responsible for extra work if he can't avoid it.	Never fails to do home work. Always willing to assume responsibility for extra work.

Rogers.[19] She attempted to analyze mathematical ability. She selected a large number of tests for the preliminary testing in an attempt to get a measure of as many forms of algebraic ability, geometrical ability, and language ability as possible. She selected as a final group of tests those that correlated most highly with the criterion of mathematical ability. Six tests were finally kept and are included in the "Rogers Test of Mathematical Ability." This test is not truly a prognostic test since some of it must be given after instruction in algebra or geometry has been given.

Another test, developed by the Bureau of Public Personnel Administration, that has combined many of the important factors of prognosis in one test is the Tests for Junior Clerks. A short description of the tests follows.[20]

Test 1. Copying and adding.
(30 items are to be selected from a much larger list of items and copied into five columns; the sum of each column is then found.)

Test 2. Business information.
(Section *A* consists of 20 questions in the multiple choice form, each with four answers from which the competitor chooses the best one. Section *B* consists of 25 statements in the completion form each of which requires the com-

[19] "Experimental Tests of Mathematical Ability and Their Prognostic Value," Teachers College Contributions to Education, No. 89, 1918.

[20] "Partially Standardized Tests for Junior Clerks," *Public Personnel Studies*, Vol. 3: pp. 346-71, 1925.

petitor to write in a word or phrase to supply missing information. Each question or statement relates to business or office transactions, mostly of a type familiar to the adult without clerical experience and about which the person with clerical aptitude is apt to have some information.)

Test 3. Alphabetical filing.
(This test calls upon the competitor to insert 20 names in the proper places in a list of 72 names in accordance with rules as to alphabetical order which are given.)

Test 4. Understanding of printed literature dealing with office problems.
(This test consists of 10 questions relating to content of a quoted section dealing with business administration or office management.)

Test 5. Social intelligence.
(This test consists of 20 questions in the multiple choice form, each with four answers from which the competitor chooses the best one. Each question deals with relations with fellow-employees or the public.)

Test 6. Understanding and following written directions.
(The competitor is asked to carry out 10 written directions relative to the checking of certain items in a table.)

Test 7. Completion of tabulations.
(In each of two tables the competitor is asked to insert 5 missing items requiring both reasoning and calculations; each table contains 8 columns and there are 5 items in each column.)

Test 8. Memory for oral directions.
(A set of directions such as might be given to a clerk are read and the competitor is asked to answer in writing 5 questions as to their meaning. In addition, a list of 15 items is read and the competitor is asked to record in writing as many as he can remember up to 10.)

Test 9. Abstract intelligence.
(One of the five alternate forms of the Army Alpha Intelligence Test is given; a minimum score of 75 on the Army Alpha scale is required.)

Test 10. Education and experience.

Another test of this type is the Hoke Prognostic Test of Stenographic Ability. This consists of 7 tests: (1) Motor

reaction, (2) Speed of handwriting, (3) Quality of handwriting, (4) Speed of reading, (5) Memory, (6) Spelling, (7) Symbols. Motor reaction is tested by having the pupils make marks in small boxes, or parallel lines with cross-sections, which they construct on the test paper. Speed in writing is measured by the number of letters written per minute. Quality of writing is measured by the Ayres Handwriting Scale. The other tests are made up of appropriate material to bring out the qualities intended to be measured.

Attempts have been made to measure mechanical ability and aptitude. Stenquist has developed two tests for this purpose. One is a performance test known as the Stenquist Assembly test. Another is a pencil and paper test of mechanical ability.

The Assembly test is generally given as an individual test although it may be given as a group test. Ten mechanical contrivances which have been taken apart are given to the pupil one at a time, and he is expected to assemble each of them. The tests come in two series. Each series is put up in a wooden box [21] with each mechanical contrivance in a separate compartment. The contents of each box are:

Series I	*Series II*
1. Cupboard catch	1. Pistol
2. Clothes pin	2. Elbow catch
3. Hunt paper clip	3. Rope coupling
4. Chain	4. Expansion nut
5. Bicycle bell	5. Sash fastener
6. Shut-off	6. Expansion rubber stopper
7. Wire stopper	7. Calipers
8. Push button	8. Defiance paper clip
9. Lock No. 1	9. Double action hinge
10. Trap	10. Lock No. 2

The directions are on the outside of the box and can be read by both pupil and instructor. Thirty minutes are allowed for the test, and additional points are given if the pupil finishes before the end of the time limit. A key for scoring is avail-

[21] Made and sold by C. H. Stoelting & Company.

able and should be used so that standard results can be obtained.

It is fair to assume that what is measured in such a test may be quite distinct from the abstract type of ability measured in other types of tests such as verbal tests. The test is fairly reliable, correlates poorly with intelligence as measured by intelligence tests and highly with achievement in trade courses.

The pencil and paper test is an attempt to provide convenient simple tests which will measure the same abilities as the Assembly test. The attempt has not been entirely successful. The tests correlate more highly with intelligence tests than does the Assembly test and not so highly with shop rank. The measure is divided into two tests. Test I is made up of 6 exercises of 3 or 4 parts each. Each part is made up of 10 pictures of things connected with mechanics, which are to be matched according to use. Test II consists of 3 exercises. Exercise 1 is made up of 19 pictures of parts of vehicles, which are to be matched. Exercises 2 and 3 consist of pictures of machines with a series of questions on each.

Another test along this line is the MacQuarrie Test for Mechanical Ability. This group performance test is designed to measure native mechanical ability. It may be used as an aid in selecting and advising pupils and other persons in activities that require manipulative skill. It consists of 7 parts, with a practice test before each test that is to be scored. Part 1 is a tracing test; Part 2, tapping; Part 3, dotting; Part 4, copying; Part 5, location; Part 6, blocks; Part 7, pursuit.

BIBLIOGRAPHY

Allen, W. S. *A Study in Latin Prognosis,* Teachers College Contributions to Education, No. 135, Columbia University, 1923.

Briggs, T. H. "Prognostic Tests of Ability to Learn Foreign Languages," *Journal of Educational Research,* Vol. 6: pp. 386-92, December, 1922.

Chambers, O. R. "Character Trait Tests and the Prognosis of College Achievement," *Journal of Abnormal and Social Psychology,* Vol. 20: pp. 303-11, October, 1925.

Clem, Orlie M. *Detailed Factors in Latin Prognosis,* Teachers College Contributions to Education, No. 144, Columbia University, 1924.

Colvin, S. S. "Educational Guidance and Tests in College," *Journal of Applied Psychology,* Vol. 5: pp. 32-38, March, 1921.

Cornell, Coxe, and Orleans. *Rating Scale for School Habits,* World Book Company, 1927.

Flemming, Cecile White. *A Detailed Analysis of Achievement in the High School,* Teachers College Contributions to Education, No. 196, Columbia University, 1925.

Flinner, I. A. "Rating Students on the Basis of Native Capacity and Accomplishment," *Educational Administration and Supervision,* Vol. 9: pp. 87-98, February, 1923.

Fuller, F. D. "Study of the Hoke Prognostic Test of Stenography," *Educational Research Bulletin of the Los Angeles City Schools,* Vol. 5: pp. 5-6, March, 1926.

Gates, A. I. and LaSalle, J. "The Relative Predictive Values of Certain Intelligence and Educational Tests Together with a Study of the Effect of Educational Achievement upon Intelligence Test Scores," *Journal of Educational Psychology,* Vol. 14: pp. 517-39, December, 1923.

Gronert, Mary Lynch. "A Prognostic Test in Typewriting," *Journal of Educational Psychology,* Vol. 16: pp. 182-85, March, 1925.

Haddock, N. P. "The Use of the Stanford Achievement Test in Indicating High School Success," *School and Society,* Vol. 25: pp. 115-16, January 22, 1927.

Henmon, V. A. C. "Prognosis of Ability in School Subjects," *Educational Research Bulletin of the Los Angeles City Schools,* Vol. 5: p. 23, April, 1926.

Herd, A. W. "Intelligence Quotient As a Prognosis of Success in Physics," *School Review,* Vol. 34: pp. 123-28, February, 1926.

Hull, Clark L. *Aptitude Testing.* World Book Company, 1928.

Johnston, J. B. "Predicting Success or Failure in College at Time of Entrance," *School and Society,* Vol. 19: pp. 772-76, June 8, 1924; Vol. 20: pp. 29-32, July 5, 1924.

Jordan, J. N. "Prognosis in Foreign Language in Secondary Schools," *School Review,* Vol. 33: pp. 541-46, September, 1925.

Kelley, T. C. *Educational Guidance: An Experimental Study in the Analysis and Prediction of High School Pupils,* Teachers College Contributions to Education, No. 71, Columbia University, 1914.

Limp, C. E. "The Use of the Regression Equation in Determining the Aptitudes of an Individual," *Journal of Educational Psychology,* Vol. 16: pp. 414-18, September, 1925.

Madsen, I. N. "The Army Intelligence Tests as a Means of Prognosis in High School," *School and Society,* Vol. 11: pp. 625-27, May 22, 1920.

Proctor, W. M. "Psychological Tests and Guidance of High School Pupils," *Journal of Educational Research,* Monograph No. 1, October, 1923.

Rogers, A. L. *Experimental Tests of Mathematical Ability and Their Prognostic Value,* Teachers College Contributions to Education, No. 89, Columbia University, 1918.

Ross, C. C. *The Relation Between Grade School Record and High School Achievement. A Study of the Diagnostic Value of Individual Record Cards,* Teachers College Contributions to Education, No. 166, Columbia University, 1925.

Rugg, H. O. "Is the Rating of Human Character Practicable?" *Journal of Educational Psychology,* Vol. 12: pp. 425-38, November, 1921: pp. 485-501, December, 1921; pp. 30-42, January, 1922; pp. 81-93, February, 1922.

Ruggles, A. M. *A Diagnostic Test of Aptitude for Clerical Office Work,* Teachers College Contributions to Education, No. 148, Columbia University.

Scates, D. E. "A Study of High School and First-Year University Grades," *School Review,* Vol. 32: pp. 182-92, December, 1924.

Somers, Grover T. *Pedagogical Prognosis,* Teachers College Contributions to Education, No. 140, Columbia University, 1923.

Stoddard, George D. "Iowa Placement Examinations," *University of Iowa Studies in Education,* Vol. 3: No. 2, August 15, 1925.

Symonds, P. M. "An Index to Studiousness," *The Pedagogical Seminary and Journal of Genetic Psychology,* Vol. 32: pp. 257-65, June, 1925.

Toops, H. A. *Tests for Vocational Guidance of Children Thirteen to Sixteen,* Teachers College Contributions to Education, No. 136, Columbia University, 1924.

Viteles, M. S. "Psychological Tests in Guidance; Their Use and Abuse," *School and Society,* Vol. 22, No. 560: pp. 350-56, September 19, 1925.

Wilkins, L. A. "Results in a Prognosis Test Given to Pupils Beginning French and Spanish," *Bulletin of High Points,* Vol. 1, No. 8: pp. 26-30, Board of Education, New York City, October, 1919.

CHAPTER XXI

New-Type Examinations

The use of standardized tests by teachers has had a pronounced effect upon the method of examining pupils. The various techniques involved in the developing of the questions of the standard test have come to be widely used in the tests and examinations of many local school systems or classes within the systems. The various steps necessary for the development of a standard test and for a good new-type examination do not differ widely. The first step in either case is to determine what is to be measured or tested. If the thing to be tested is the pupil's knowledge of the various phases of *Ivanhoe*—the characters, the narrative, the historical setting, the customs of the people, the author, or any other outcome expected from the study of it—then this becomes the thing to be measured or tested.

The second thing to determine in this instance would be exactly what characters in this story the child should be expected to know and what he should be expected to know about them. In a similar manner each of the points of expected knowledge should be analyzed. The new-type or objective examination may differ from the standard test in this respect. In developing the new-type or objective examination for a local school system or for a class in the system, the expected knowledge outcomes will probably be determined by the local course of study, by local teachers working together, or by a single teacher who is teaching *Ivanhoe*. In the development of a standard test, on the other hand, many well chosen courses of study would be analyzed, or teachers in many widely distributed schools including *Ivanhoe* in the work of a given

grade would list the important knowledge outcomes expected from its study. From these various sources the makers of a standard test would determine those items that should be included.

The third step in developing a new-type examination is to determine the best method of testing for these expected knowledge outcomes of the study of *Ivanhoe*. For example, what type of question is best suited to test the child's knowledge of the approximate date of the setting of the story? To answer this question a knowledge of the various types of questions or testing techniques available would be required. These types are discussed on pages 484-497. At the time the test is made out, the answer sheet or key should be developed. This key should include the responses that are to be counted as correct and in many cases examples of responses nearly correct that are to be counted as wrong.

The fourth step is editing the new-type or objective examination with proper statements of the new-type questions in accordance with such rules as are given in this chapter.

The fifth step is determining the rules for the giving of the examination and for the scoring of it. Is the whole test to be given with or without a time limit? Is the pupil to be allowed the use of any supplementary materials? Is the pupil allowed to ask questions? How are various parts of the test to be scored? How are the scores to be interpreted? How are they to be transmitted into marks? What analysis of the results of the test might be made for purposes of more effective learning by the pupil?

The sixth step is the giving of the test under the conditions set out under the fifth step.

The seventh step is scoring the test with the keys developed under step three by the rules of scoring determined under step five. The scores should then be interpreted or transmitted into percentiles or marks.

The eighth and last step may be by far the most important

in its usefulness. It is the diagnostic analysis of the results of the test by the methods already described in the discussion dealing with standard tests in previous chapters.

In developing the standard test, one would experimentally try out each of the steps just stated. Decisions made concerning procedure would be based upon experimental data obtained in trying out the material developed. In developing the new-type examination, however, one has to make a great many arbitrary decisions based on his best judgment and the results of experimentation with other new-type tests. For example, it has been found to be better practice to place the easier questions first in the list of questions. In the development of a standard test this information can be experimentally gained by giving the tryout form of the test to a large number of pupils in different schools. Based on the results of the experimental tryout, the easiest question will be the one missed by the smallest number of pupils. However, in the development of a new-type examination for use the first time with a class, it is not definitely known which of the questions are the easiest for the pupil to answer. It will be necessary for the maker to decide arbitrarily according to his best judgment what question should come first.

In the interpretation of results, the difference between standard tests and informal objective tests is marked. In the latter, the results of only one class or one school system are available to determine the significance of the results; whereas in the former, comparable data are available from many other sources.

The steps that are included and the scientific procedure that is necessary in the development of a standard test are given in the next chapter. The discussion here is intended only to point out the likeness in development between new-type or objective tests and the standard test. This likeness is brought to our attention because it is sometimes felt that these two forms of testing are in conflict in content, in method, and in purpose.

There is much likeness in both content and form or method. Nevertheless, each type of test has its special use and value, and both, taken together, may serve complementary and supplementary purposes in education.

The three types of examinations—traditional essay type, new objective type, and standard test—have been discussed in Chapter II. There is a specific place for each of these types of examinations. It is to be expected that the objective type test will supplant the essay test in a great many of the examinations to be given in the future. It is equally true, however, that the essay-type test will remain most useful in those phases of school work where extensive discussion is desired and where objective type questions will not adequately measure organization and expression of thought.

CHARACTERISTICS OF NEW-TYPE OR OBJECTIVE EXAMINATIONS

The various types of objective tests or questions to be used in the test may be classified as follows.

A. Recall Types
   1. Complete recall (simple recall, enumeration)
   2. Completion
B. Recognition Types
   1. Alternate response (true-false, yes-no, right-wrong, same-opposite, select the correct word)
   2. Multiple response (single choice, plural choice, best answers)
   3. Matching
   4. Identification
   5. Rearrangement
   6. Correction

RECALL TYPES

The *simple recall* question has been in use in examinations for some time. "When did Columbus discover America?" is a question of this type. Such questions have been found in

examinations almost since written examinations began. The only change that has been made in it for its use in the new-type examination has been in its form of statement. It would now appear in a new-type examination: "In what year did Columbus discover America?....." The improvement facilitates scoring and makes possible more extensive use.

The recall type of question is also used in what is sometimes called the enumeration question. "Make a list of three widely used food products that the United States imports from Central and South America." 1.....

2.....

3.....

The recall type of question has certain advantages and disadvantages.

Advantages

1. It is like the question often used in the classroom.
2. It does not suggest possible answers and thus make guessing possible.
3. It allows some freedom of expression.
4. The enumeration type is adapted to some extent to thought questions.
5. It is usually more easily scored than the essay type.
6. It is the type of question most valuable to use when a test of total recall is desired.

Disadvantages or Limitations

1. Because an examination made up of this type of question is not purely objective and because the answers are not controlled,

(a) It requires longer for the pupil to take it than is the case with some other types of examinations covering the same ground;

(b) It is not easily scored, and the scoring is not likely to be the same with two teachers scoring the same test.

2. In the case of many facts and principles in school achievement total recall is not desired. A few dates of history may be of such a type that a pupil should be able to recall them exactly. However, it is more desirable that the child be able to recognize certain other dates in their association with certain events. Where the total recall is not a desirable end of instruction, the recall type of question is not adapted and should not be used.

Following is an example of a *completion* exercise.[1]

Silas Marner

....., chose the night of a New Year's party at Red House, the home of Squire ....., to reveal to the squire that she was the ..... of his eldest son. Due to the effects of a ..... she had taken and due to the weight of the ..... that she was carrying she fell exhausted in the snow. ....., in the meantime, had opened the door of his cottage to see if there might be any trace of his missing ..... While he was gazing out into the night an attack of catalepsy came upon him. The ..... which streamed out through the open door attracted the ....., who had wakened, and led her to the cottage. Unnoticed by ....., she entered and toddled over to the warm ....., where she soon fell asleep. When the weaver recovered consciousness and discovered the yellow-haired ..... on his ....., he thought his ..... had been returned.

The two important things to note about this exercise are that (a) the knowledge to be tested for is omitted, (b) no other correct word could be used to fill the blank. These two important rules should be observed in the construction of completion tests. Failure to observe the two rules will make the completion test relatively worthless as a measure of achievement. Another danger involved in its use is that too many

[1] Through the courtesy of Dr. C. G. F. Franzen, Professor of Education, Indiana University.

blanks may be used in any one sentence or item; thus the test will become a puzzle rather than a test of achievement. Take for example the first sentence in the exercise.

...., chose the ..... of a ..... party at ....., the ..... of Squire ..... to ..... to the ..... that ..... was the ..... of his eldest .....

In this case the completion test becomes a puzzle, encourages pure guessing, and is no longer valuable as a test.

One very good rule that may be applied to simple recall, enumeration, and completion questions is this. Determine what question is to be answered, write out the answer, and then shape the question in such a way as to test the pupil's knowledge of the correct answer. It may be desirable, for instance, to test the pupil's knowledge of the declarative sentence. The question is, "What is a declarative sentence?" The answer is, "A declarative sentence is one that makes a statement." A simple recall question could be stated: "What kind of sentence makes a statement?" A completion question could be made thus: "A ..... sentence is one that makes a statement," or "A sentence that makes a statement is a ..... sentence."

The completion sentence may be used in the single statement, but probably finds its best use in a more complex exercise like the one on *Silas Marner.*

The limitation of this type of examination is that it is difficult to construct properly and to score fairly. The usual completion test has more than one possible correct answer if it goes beyond the testing of facts.

The completion type of question can well be used to advantage in thought questions if the questions are properly constructed. This type of question allows for testing of continuous thought along a certain line. An example of this is seen in the completion exercise on *Silas Marner.*

RECOGNITION TYPES

Of the recognition types of questions or examinations one of the most commonly used is the *alternate response*. This type of question has several forms, which are illustrated.

True False

"The Missouri Compromise provided that Maine be admitted as a slave state." True False

Yes No

"Did the Missouri Compromise provide that Maine be admitted as a slave state?" Yes No

Right Wrong

"Mother knowed it all the time." Right Wrong
Correct Incorrect

Cross out the Incorrect Word

"Mother (knew ~~knowed~~) it all the time."

Same Opposite

Old Young Same Opposite

Not all of the various forms of the alternate response type can be used in the same way. The true-false question is much used. This type of question (a) is applicable to many school subjects, (b) is rather easily constructed, (c) can be made to sample widely the content of any subject in a short time, (d) can be rapidly scored in an objective manner, (e) can range from fact questions that require only recognition to thought questions that require reasoning.

Three major objections have been made to it. The first objection is that false statements or untruths should not be made to children in such a straightforward and bold fashion

as they are in mimeographed and printed pages. To some this is psychologically objectionable. Those who put forth this objection feel that it applies only when such tests are given to children of elementary and junior high school level. No proof has been given to indicate that such a feeling is warranted. If the child were reading the statements without first being told "some of the sentences below are true, some of them are false and you are to decide which ones are true and which ones are false," the criticism would seem to be more nearly valid. However, where such a warning is given, it seems extremely doubtful whether a negative result is obtained. On the other hand, it might be argued with no more proof that practice in distinguishing between true and false statements concerning school content would develop the habit of critically distinguishing between the true and false statements of after-school life. In the true-false type of test there is no proof of harm, the child is not exposed to falsehood for any length of time, and he is warned that "some of them are false." While this form of testing is different from the usual form, it is not necessarily psychologically objectionable.

The second difficulty in connection with this type of test arises because of the effect of guessing. In any alternate response type of question this same objection holds true. If one hundred true-false statements or other form of alternate response questions were given in Icelandic language, and the person taking the test knew nothing of the language and was told to indicate the statements as either true or false, it is probable that he would get half of them right. Because of the law of chance or the fifty-fifty chance of guessing correctly, fifty of the statements would be correctly marked and fifty incorrectly marked. It is obvious that the person taking the test would achieve this score of 50 merely by guessing; therefore the test has tested nothing. In another case suppose one hundred true-false statements in English were given on a subject about which the pupil has studied. The pupil has a

positive knowledge concerning fifty of them and marks them accordingly. Concerning the remainder he is not sure, but marks them. According to the law of chance, if he knew nothing concerning the last fifty, he would still get approximately twenty-five right by guessing. This gives the pupil seventy-five correctly marked and twenty-five incorrectly marked. It would seem fair to mark the score on the test according to the formula No. Right minus No. Wrong. In the first case this would give a $50 - 50 = 0$, the score. In the second case $75 - 25 = 50$, the score obtained because of true knowledge of the first fifty. However, in the second case, if it is like most tests or examinations, the pupils know something about many of the last fifty statements. The question really is how many of the statements are marked by pure guessing and how many are marked from a fringe of knowledge concerning the statement.

The third difficulty encountered in the use of the true-false statement is that of ambiguity of statement. This is an important difficulty, which applies to all types of examinations and even to legal and literary expression. This difficulty is one to which true-false statements are peculiarly susceptible, however, because the way in which the fact or principle is stated may determine whether it is true or false. The statement is isolated and can have only its own meaning, without explanation or interpretation. If the statement were in a paragraph following a discussion, it would be more easily judged. This one difficulty of ambiguity more than any other one thing hinders the effectiveness of the true-false examination. Weidemann [2] has discussed this phase in detail. Under sentence structure alone are listed twenty-two rules with illustrations of "Given" and "Improved" statements. The following will illustrate.

[2] *How to Construct the True-False Examination,* Teachers College Contributions to Education, No. 225, Columbia University, 1926.

"Try to place modifiers (clauses, phrases, and words) near the word or words they actually modify.

"Given (Ambiguous): The young child reads words almost as readily inverted as upright."

Do the words *inverted* and *upright* refer to the *child* or to the *words* in the statement? The statement may mean that the young child can read almost as readily resting on his head as standing on his feet. That may be true for some young children and false for others. The statement may mean that the young child can read inverted words almost as readily as upright words. This should be either true or false in terms of reliable evidence.

"Improved: The young child reads inverted words almost as readily as he reads upright words."

The following rules may be helpful in the development and use of the true-false test. (They can also be applied to the yes-no type of test.)

1. Make the statements as clear, concise, and definite as possible.

   The simple declarative sentence should be used where possible. The test should be in the vocabulary of the pupils of the given grade.

   It is well to have other people read the statement to see if they understand it to mean what *you* intend it to mean. Such a procedure will also give an opportunity for correction of errors of punctuation and spelling.
2. Include only statements that are consistently either true or false. If the question is a debatable one but included in a book or text being studied, it should be stated that the correctness of the statement would be judged by the text. Many phases of subject matter are neither absolutely true nor absolutely false. Such subject matter should be tested with some other type of question.

3. Do not include merely negative words or catch words in a positive sentence to make it a false sentence.

   Julius Cæsar was not a Roman.
   Theodore Roosevelt was president of the United Kingdom.

4. Approximately one-half of the statements should be true and approximately one-half false. The true and false statements should follow each other in no regular order.
5. A test of the true-false type should include at least fifty statements. A test of a hundred statements is more desirable.
6. The test should be preceded by the statement of directions for taking, which should include *Do not guess.*
7. The best method of indicating whether the statement is true or false is by encircling either the T or F before the statement or marking true statements + and false statements o.
8. The best method of scoring has not been determined. It seems that the most common practice is number right minus number wrong (minus scores counted as zero). It is doubtful, however, whether such corrections materially increase the value of the measuring instrument.

Other types of alternate response questions have been largely used in tests of language usage. Such tests are of an alternate response type that is peculiarly adapted to this subject.

The *multiple response* type of question is not only widely used but is the least criticized of any of the objective type questions. Examples of the various forms of multiple response are given.

Multiple-response, single-choice

The greatest wheat-producing country of South America is (1) Brazil, (2) Peru, (3) Argentina, (4) Ecuador, (5) Chile.

Multiple-response, plural choice

Complete the following sentences by putting a cross (X) before *three* items that complete each statement to make it true. Remember to put the (X) before only *three* items.

39-40-41. Great Britain probably would not be able to support her present population except that
- a. she is a great military nation
- b. she is a great shipping nation
- c. she is a great manufacturing nation
- d. she has a great amount of natural water power
- e. she has a splendid tropical climate
- f. she has important coal and iron deposits
- g. she has a large area of virgin forests [3]

Best Answer

Why should forests in the United States be conserved?
1. They increase the supply of oxygen.
2. They make excellent places for hunting game.
3. They supply building materials and decrease floods.
4. They enrich the soil and shade the ground.[4]

There are two possible difficulties with this type of test. In giving the various responses or best answers, there is danger of having more than one answer that may be correct or so nearly correct that it would be correct to mark either one right. It is difficult to make the wrong responses or answers plausible and seductive and still wrong beyond question. The only way to overcome this difficulty is to work with other teachers in the preparation of such questions or to have other teachers make an independent check of the responses to eradicate ambiguity, to check the correct response, and to suggest better and more plausible wrong answers. It is also a good practice, after giving a test, to go over with the pupils the reactions that they

[3] Indiana Composite Achievement Test, Form *B*, Bureau of Coöperative Research, Indiana University.

[4] B. R. Buckingham and P. R. Stevenson, "Geography Test, United States Information Problems," Public School Publishing Co.

had to questions in order to determine if the reactions were of the type desirable. In this manner any particular test may be improved for future use and in addition the experience will serve as a suggestion of the difficulties that grow out of the faulty preparation of tests. Such methods of checking cannot be too strongly urged in the preparation of any of the new-type questions.

Many questions do not have enough plausible answers to make a good multiple response question. Example:

"The selling price of corn in November compared with the price in May is usually (higher, lower, approximately the same)."

It is desirable to have at least four or five responses in a multiple response question. To add more to the example would not add anything to the question because the pupil would still make his judgment on the basis of the first three responses. These same criticisms or difficulties apply to multiple response-single choice, to multiple response-plural choice, and to best answers types of questions.

While the multiple response-single choice and multiple response-plural choice statements are likely to lend themselves best to factual type questions, nevertheless they can be made to function in such a way as to cause the pupil to use judgment and reasoning in determining the answer. The multiple-response type of question is easier to prepare than the completion or true-false type. It can be made easy to score and is entirely objective. If five to seven responses are given, the element of guessing is almost eradicated. The multiple-response type of question is valuable in that it measures what one can recognize, which is a much wider field than what one can recall.

The best-answers type of question is highly desirable for thought and judgment questions. While such questions require much space in mimeographing and typing, there is no new-type

question of more value if the examination of the problem type is desired. If this type of question is carried into a plural choice, it becomes what is called an *organization* type of question. An example of such a type is given.

TEST I

Organization Test[5]

Number Right ..... × 2 = ... Score

Place the letter (X) before the ten statements which show steps in America's growth toward democracy.

.... 1. Increased representation of the lower classes in colonial assemblies.
.... 2. Granting of religious freedom in Maryland and Rhode Island.
.... 3. Securing of independence from Great Britain.
.... 4. Hamilton's plan of establishing credit for the new government.
.... 5. Increased power given to the central government by the Federalist Party.
.... 6. The purchase of Louisiana by Thomas Jefferson.
.... 7. Increased power given to the Supreme Court by the decisions of John Marshall.
.... 8. Growth of nationalism following the War of 1812.
.... 9. Struggle to do away with property and religious qualifications for voting.
....10. Demand of the laboring class for the support of the public schools by taxation.
....11. Demand of the manufacturing class for a protective tariff.
....12. The use of the Spoils System in appointing men to government positions.
....13. The reconstruction of the southern states after the Civil War.
....14. The freeing of the slaves.
....15. Reclaiming of waste lands by the United States government.

[5] Denver Public Schools Curriculum Test in Social Science, Grade 8A.

....16. Direct election of senators by the people.
....17. Taxing of large incomes.
....18. Granting of the voting privilege to women.
....19. Government restriction on the manufacture and sale of intoxicating liquors.
....20. Present demand for child labor laws.

The *Matching* type of question is valuable in measuring ability to associate facts or ideas; for example, ability to associate books and authors, dates and events, or two statements of the same idea. An example follows.[6]

In the blanks before the column at the right place the number of the author which corresponds to the list in the column at the left. The first one is marked as it should be.

93. Holmes	97a Thanatopsis
94. Lowell	....The Spy
95. Wordsworth	....Christmas Carol
96. Cooper	....The Man Without a Country
97. Irving	....Daffodils
97a. Bryant	....Legend of Sleepy Hollow
98. Hale	....Chambered Nautilus
99. Dickens	....Vision of Sir Launfal

This type of question is of value (a) where association test is desired, (b) where the number of items is large enough to prevent guessing. The test is completely objective and easily scored.

The *Identification* test is not so very different from the Matching test. It is best used in the cases where it is desirable to test the ability to apply a principle to concrete illustrations by identifying such situations as examples of the principle. This ability may be tested by giving one principle and asking for identification to be made from examples given, by giving one concrete illustration and having it identified from

[6] Indiana Composite Achievement Test, Form *A*, The Bureau of Cooperative Research, The School of Education, Indiana University.

several principles, or by listing principles and concrete illustrations and having them matched as in the matching exercises. Following is an example.

"Below are several lists of the food for the breakfast of a child of 7 to 9 years of age. From what you have learned from your study of balanced meals for children, select the one that you think would be the most suitable meal." Several menus are then given from which to choose.

This same type of test may be used in a map study in such a question as this: "From the map select a city that is the national capital whose altitude is more than 5,000 feet. Write the name here ....." This requires identification of the markings of altitude and of capital cities on the map. While this type of test is somewhat limited in use to certain kinds of subject matter, it is valuable in bringing out certain types of thinking in a comprehensive test.

The *Rearrangement* test is of most value in testing knowledge of chronological order or in logical steps of procedure in a type of work where the order of the steps is important. The example is one taken from Indiana history.[7]

In each of the lists below are the names of four events. Arrange these events in the time order in which they occurred by putting an (a) before the event that occurred first, a (b) before the event that occurred second, a (c) before the event that occurred third, and a (d) before the event that occurred last. The first one is marked as it should be.

a Indiana becomes a territory	..Founding of Vincennes
d Election of first state officers	..Founding of Indianapolis as capital
c Making of Constitution	..Founding of New Harmony
b Application for admission into Union	..Founding of Corydon as capital

[7] *Ibid.*

The following general rules are of importance in making a new-type or objective examination. Some of these have been stated before but are here summarized:

1. Determine what is to be tested before beginning to develop the examination.
2. Cover every phase of the subject or state what phases are being covered. This rule applies both to the scope of the work covered and to the outcomes of instruction expected. For example, "This test covers the important facts that children are expected to learn in the period of American History from Discovery and Exploration to the Civil War."
3. Take great care to avoid ambiguous questions and statements. Since ambiguity is an especial difficulty with objective type tests, methods of overcoming it must be used. It is suggested that such tests be checked for ambiguity by other teachers. Other things being equal, a short question is desirable.
4. The questions which are assumed to be pertinent to the subject should range in difficulty from very easy to very difficult. The easier questions should come first.
5. There should be a large number of questions in an examination. Other things being equal, this will increase the reliability of the examination. Since the time that is required in answering any single question is very short, it is possible to give a very much larger number of questions in the new-type objective examination than can be included in the traditional essay-type examination. One should learn to think in terms of one hundred questions instead of in terms of the traditional ten.
6. Each question should be an independent unit. Neither its statement nor answer should in any way depend on any other question.

7. Use as many types of questions as will well fit into the examination content that is desired. This is suggested because it is generally true that different types of thinking are required to answer the various types of questions. This variety in types of questions will on the whole give a more comprehensive measure of the pupil's ability and achievement.
8. It is well to arrange questions somewhat in sequence according to the way in which the pupils have learned the content of the course. This should not be done, however, if it conflicts with other rules given.
9. In the development of the examination, try wherever possible to prevent or reduce the guessing element. In multiple-response types use many responses to prevent chance guessing. There should be at least five responses in a single choice, and at least one more than twice the number of correct responses in plural choice. In placing the correct response in any type of question, there should be no fixed order of arrangement. If the second response is most always the correct one, this arrangement will tend to develop guessing and ruin the results of the tests. Chance order should be used in all types of new-type questions wherever possible.
10. Since a general weakness of new-type examinations is that they tend to be too factual, a special effort should be made to see that they include enough judgment and reasoning questions to bring out all phases of the subject.
11. Make out the key or expected correct answers at the same time that the test is made out.
12. Make careful directions for each part of the test. There should be general directions at the beginning of the test and specific directions for each part of the test, with an example to show what is to be done if this will make the question clearer to the pupil. Remember that in these examinations you are testing, not to find out if the pupil

knows how to take a test, but to find out what he knows about the content of the test.

13. In general it is better not to weight the questions. The best method, when both accuracy and ease of use are considered, is to consider one part of the examination as important as any other. The correction of the score to correct chance error or guessing is not weighting.
14. The test should be duplicated so that each pupil can have a copy.
15. When the test is scored, it is best to make a distribution of the scores, group the distribution, and turn the scores into school marks. Example:

Group Scores Made on History Test	*No. Pupils*	*Mark*
80-99	2	A
60-79	6	B
40-59	14	C
20-39	7	D
0-19	3	F
Total pupils	32	

16. Make diagnostic analysis of pupil errors.
17. Prepare and use remedial work.

Following are illustrations of the use of some of the various types of questions that have been discussed in this chapter.

True-False Test

Directions: Examine each statement below and decide whether it is true or false. If the statement is true, place a "T" after it on the dotted line; if the statement is false, place an "F" after it on the dotted line. Do not *guess*. You have thirty minutes for Test 1.

1. H. G. Wells showed considerable critical acumen in his extensive examination of source material which preceded his writing of the *Outline of History*. F

2. Reviewing is frequently done by those who have not read the books they review. T
3. The aim of the biographer should be totality. T
4. A librarian should primarily select books for the benefit of educated persons. F
5. Tennyson's life written by his son illustrates the fact that a member of the family can write a better biography than can some outsider. F
6. Reviews in publishers' announcements are very reliable. F
7. Neither the whitewashing method nor the method of suppression is justified in biography. T
8. Frequent articles on the use of books in schools and libraries appear in *Publishers' Weekly*. T
9. The folio is the best size for the Library for general reading. F
10. J. A. Froude was often beset by vagaries of memory. T
11. Topicality is an essential quality of a good annotation. T
12. George Borrow's *Lavengro* and *Romany Rye* are largely fictitious in character. F
13. For long-continued reading, books printed in 11-point are most satisfactory. T
14. Many series of children's books should be bought for the library. F
15. Straightforward news of books is better than half-baked criticism. T
16. The principles which define a library's policy in book selection are based on the format of the book. F
17. Material on holidays has been listed in the Children's Catalog by analyzing many children's books of stories and poems. T
18. In his *Life of Cromwell*, Morley chooses the dramatic rather than the philosophical method.[8] T

[8] American Library Association, Summer Library Institute, 1926, held at University of Chicago: New-Type Objective Examination Questions.

Multiple Response-Single Choice

Taken from Objective Examination on "How to Use the Library." (Answers underlined are those given as correct.)

The articles of Confederation can be found in
1. New International Year-Book.
2. Statesman's Year-Book.
3. Harper's Encyclopedia of U. S. History.
4. Hart & McLaughlin-Cyclopedia of American Government.
5. World Almanac.

Sketches of living Americans can be found in
1. National Dictionary of Biography.
2. Lippincott's Biographical dictionary.
3. Who's Who in America.
4. Appleton's Cyclopedia of American biography.
5. Century Cyclopedia of names.

The origin of famous names in fiction may be found in
1. Baker's Guide to best fiction.
2. Brewer's Readers hand-book.
3. Statesman's Year-Book.
4. Readers' Guide.
5. World Almanac.

The author and source of poetry and recitations may be found in
1. Firkins—Index to short stories.
2. Granger—Index to poetry and recitations.
3. Book Review digest.
4. Ward—English poets.
5. A. L. A.—Index to general literature.

The leading articles in all the important periodicals are indexed in the
1. Cumulative Book index.
2. Book Review digest.
3. Reader's Guide to Periodical Literature.
4. Card Catalog.
5. Severance—Guide to Periodicals.

The Statistical Abstract of the U. S. contains

1. Statistical tables of the last census.
2. Abstract of deeds.
3. Current statistics.
4. Statistics about government.
5. Statistical maps of the last census.[9]

Completion Test

First Year Latin Test on Case.

This part of the test has to do with rules for the use of case forms. Fill in the blanks with the correct case, e.g., the direct object of a verb is in the *accusative* case. (Accusative is the word to use in the blank.)

1. The indirect object of a verb is in the ..... case.
2. The subject of a finite verb is in the ..... case.
3. The ..... case is used to denote the person addressed.
4. A predicate noun used with a finite form of **sum** is in the ..... case.
5. With names of nouns and small islands, also with **domus,** the place where some act occurs or something exists is denoted by the ..... case.
6. A noun used to tell how long an act or situation continues is put in the ..... case.
7. Verbs meaning *to favor, please, trust, obey, serve, resist, envy, threaten, pardon, spare,* and *persuade,* take their object in the ..... case.
8. With passive verbs, the noun or pronoun which indicates the person by whom an act is done is put in the ..... case with **a** or **ab.**
9. A noun used to tell extent in space is put in the .....
10. With the future passive participle the ..... case is used to indicate the person by whom an act must be done or ought to be done.

Multiple Response-Single Choice Test

Directions: Draw a line under the word, or group of words, in parentheses, which, in a bookkeeping sense, most nearly means the same as the first word in each line.

[9] American Library Association, Summer Library Institute, 1926, held at University of Chicago: New-Type Objective Examination Questions.

1. Liability—(mortgage, debt, note, bond).
2. Insolvent—(saving, bankrupt, investment, liability).
3. Trial Balance—(journal entries, list of ledger accounts, balance sheet, assets).
4. Asset—(cash, property, wealth, profits).
5. Inventory—(sales, stock, income, investment).
6. Note—(check, written promise to pay money, cash, draft).
7. Revenue—(wealth, taxes, protective tariff, income).
8. Merchandise—(profits, goods, bond, money).

Multiple Response-Plural Choice Test

Complete the following sentences by putting a cross before two items that complete the statement to make it true. The example is marked as it should be. Remember you are to put the cross (X) before only two items.

Example: Indiana ordinarily raises

X a. corn
b. cotton
c. hemp
d. coffee
X e. wheat

Two articles extensively manufactured in Indiana are

a. wagons
b. woolen cloth
c. shoes
d. watches
e. brick

Practically every river of any size has

a. waterfalls
b. a canyon
c. a basin
d. a delta
e. a good harbor

In the northern part of Asia are

a. the tundras
b. the savannas
c. the campos
d. the pampas
e. the steppes

Important rivers found in South America are
a. the Danube
b. the Congo
c. the Amazon
d. the Niger
e. the Plata

The inclination of the earth's axis to the plane of its orbit causes
a. day and night
b. the different seasons
c. the variation in rainfall
d. the ocean tides
e. the variation in length of day and night.

Classification[10]

Directions: Place the number before the age group given below after the title in the following list of books to indicate the age of the child to whose comprehension the book is best suited:

1. From three to six years.
2. From six to ten years.
3. From ten to fifteen years.

Potter	Peter Rabbit	1
Lorenzini	Pinocchio	2
Masefield	Jim Davis	3
Alden	Moral Pirates	2
Altsheler	Free Rangers	3
Spyri	Heidi	2
Lewis	Friendly Adventures of Ollie Ostrich	1
Munroe	Flamingo Feather	3
Burnett	Sara Crewe	2
Lofting	Dr. Dolittle	2
Clark	Poppy-seed Cakes	1

[10] American Library Association, Summer Library Institute, 1926, held at University of Chicago: New-Type Objective Examination Questions.

Waterloo	Story of Ab	3
MacDonald	At the Back of the North Wind	2
Grimm	Fairy Tales	2
Bannerman	Little Black Sambo	1
Grahame	Wind in the Willows	2
Pyle	Men of Iron	3
Parry	Arabian Nights	2
Defoe	Robinson Crusoe	2
Kipling	Captains Courageous	3

Best Answer[11]

1. Classification of books by subject in a library is desirable because
   - It keeps books of like size together.
   - It produces an orderly arrangement by subject.
   - It allows for expansion on the shelves.
   - It makes the shelves look tidy.
   - It saves space.

2. The process of classification requires only
   - A knowledge of the subject matter of the book in connection with the practice of the library.
   - A clear understanding of the author's object as explained in the preface.
   - A careful reading of the whole title-page.
   - Reading of extensive book-reviews.
   - Reading of the table of contents

3. Library of Congress has made its section of Social Sciences larger than the other sections because
   - This subject is more important than others.
   - It has been assumed that this section will be unusually large in the Library of Congress.
   - More books in this field will be published than in other fields.
   - This class is very easily divided.
   - The majority of people are interested in this subject.

[11] American Library Association, Summer Library Institute, 1926, held at University of Chicago: New-Type Objective Examination Questions.

4. The relative index of the D C may best be used
   In determining the subject heading for a book.
   In determining what the book is about.
   In determining synonymous meanings of words.
   In determining the length of the class number.
   In determining the relative position of a certain subject.

Organization Test [12]

Instructions to be read aloud by teachers and pupils together.

This test is given to see how well you are able to arrange groups of words into sentences. . . . These words, as they are printed, do not make sense. How many see how they could be arranged to make a good sentence? . . . All you need to do is to write the numbers in the proper order in the box after each exercise. Write them in the box after the exercise.

	Write the Numbers in These Spaces
1. (1) a dog, (2) a boy, (3) had.....	2 3 1
2. (1) of the cold, (2) afraid, (3) they were.....	3 2 1
3. (1) I am, (2) see, (3) how tall.....	2 3 1
4. (1) heard, (2) the fly, (3) they, (4) buzz.....	
5. (1) wanted, (2) to go home, (3) him, (4) the dog.....	
6. (1) and made them, (2) into the palace, (3) their own, (4) took them, (5) they.....	
7. (1) go home, (2) the goat, (3) would not, (4) cried, (5) the boy, (6) because	

[12] Devised by Harry A. Greene, State University of Iowa, 1918.

(1) (2) (3) (4)
8. of rejoicing, to have, all wanted, they were,
(5) (6) (7)
a great feat, that they, and so thankful,
(8)
so happy ..... [ ]

(1) (2)
9. in England, of the first Thanksgiving Day,
(3) (4) (5) (6)
the ocean, begins, the story, far across..... [ ]

(1) (2) (3)
10. his brother, Washington, in a railway coach,
(4) (5) (6)
when, he did not ride, went to visit..... [ ]

Identification Test

Read these story summaries, then indicate the title of story, and author, and nationality of the author:

Story Summary	*Title*	*Author*	*Nationality*
1. Period—the Middle Ages; characters—a gallant young soldier, a fair maiden in distress, a cruel relative; the setting of the story—a castle with a trap door and men-at-arms behind the arras; the plot—a girl compelled to see a gallant youth hanged or save him by marrying him.			

BIBLIOGRAPHY

Batson, William H. "Reliability of the True-False Form of Examination," *Educational Administration and Supervision,* Vol. 10: pp. 95-103, February, 1924.

Black, N. H. "Some New and Old Types of Physics Tests," *School Science and Mathematics,* Vol. 26: pp. 272-74, March, 1926.

Brinkley, S. G. "Values of New-Type Examinations in High School with Special Reference to History," *Journal of Educational Research,* Vol. 13: pp. 64-65, January, 1926.

Brinkley, S. G. "New-Type Examinations in the High School," Teachers College Contributions, No. 161, Columbia University, 1924.

Brown, C. M. "Construction and Use of Information Tests in Home Economics," *Journal of Home Economics,* Vol. 16: pp. 251-56, May, 1924.

Buckingham, B. R. *Research for Teachers,* Chap. 6: pp. 221-253, Silver, Burdett and Company, 1926.

Chapman, J. C. and Toops, H. A. "A Written Trade Test: Multiple-Choice Method," *Journal of Applied Psychology,* Vol. 3: pp. 358-65, April, 1919.

Crawford, C. C. and Raynaldo, D. A. "Some Experimental Comparisons of True-False Tests and Traditional Examinations," *School Review,* Vol. 33: pp. 698-706, November, 1925.

Dickinson, Z. C. "Suggestions Toward Improving Examination Marks," *Bulletin, University of Minnesota,* Vol. 26: pp. 31-36, August, 1923.

Gates, Arthur I. "The True-False Test as a Measure of Achievement in College Courses," *Journal of Educational Psychology,* Vol. 12: pp. 276-87, May, 1921.

Gray, W. S. "The Value of Informal Tests of Reading Accomplishment," *Journal of Educational Research,* Vol. 1: pp. 103-11, January, 1920.

James, B. B. "The Modern Test," *School and Society,* Vol. 19: pp. 209-13, February, 1924.

Johnson, George R. (Director). "Objective Achievement Tests Constructed and Used in St. Louis," *Public School Messenger,* Vol. 25, November, 1927, St. Louis Public Schools.

Laird, Donald A. "A Comparison of the Essay and the Objective Type Examinations," *Journal of Educational Psychology,* Vol. 14: pp. 123-24, February, 1923.

McCall, W. A. "A New Kind of School Examination," *Journal of Educational Research,* Vol. 1: pp. 33-46, January, 1920.

McClusky, H. Y. and Curtis, F. D. "Modified Form of True-False Examination," *Journal of Educational Research,* Vol. 14: pp. 213-24, October, 1926.

Miller, George F. "A Variation in True and False Achievement Test," *School and Society,* Vol. 20: pp. 250-51, August 23, 1924.

Monroe, W. S. "Written Examinations Versus Standard Tests," *School Review,* Vol. 32: pp. 253-65, April, 1924.

Monroe, W. S. "Written Examinations and Their Improvement," *Bulletin No. 9,* Bureau of Educational Research, University of Illinois, 1922.

Odell, C. W. *Traditional Examinations and New-Type Tests.* The Century Company, 1928.

Odell, C. W. "Another Criticism of Tests Requiring Alternative Responses," *Journal of Educational Research,* Vol. 7: pp. 326-30, January, 1923.

Orleans, Jacob S. and Sealy, Glenn A. *Objective tests.* World Book Company, 1928.

Paterson, D. G. *Preparation and Use of New Type Examinations,* World Book Company, 1925.

Paterson, D. G. "Improving the Examination Function in Teaching," *Bulletin,* University of Minnesota, Vol. 26: pp. 47-67, August, 1923.

Remmers, H. H. and E. M. "Negative Suggestion Effect on True-False Examination Question," *Journal of Educational Psychology,* Vol. 17: pp. 52-56, January, 1926.

Roback, A. A. "Subjective Tests Versus Objective Tests," *Journal of Educational Psychology,* Vol. 12: pp. 439-44, November, 1921.

Ruch, G. M. *The Improvement of the Written Examination,* Scott Foresman & Co., 1924.

Ruch, G. M. and Stoddard, G. D. "Comparative Reliability of Five Types of Objective Examinations," *Journal of Educational Psychology,* Vol. 16: pp. 89-103, February, 1925.

Ruch, G. M. and Stoddard, G. D. *Tests and Measurements in High School Instruction,* Chapters 15-16, World Book Co., 1927.

Ruch, G. M. and DeGraff, H. H. "Corrections for Chance and Guess Versus Do Not Guess Instructions in Multiple Response Tests," *Journal of Educational Psychology,* Vol. 17: pp. 368-75, September, 1926.

Ruch, G. M. *Objective Examination Methods in the Social Studies,* Scott, Foresman & Co., 1926.

Russell, Charles. *Classroom Tests,* Ginn & Company, 1926.

Spencer, P. L. "Improvement in Teaching by Means of Home-made Non-Standard Diagnostic Tests and Remedial Instruction," *School Review,* Vol. 31: pp. 276-81, April, 1923.

Tharp, James B. "The New Examination Versus the Old in Foreign Languages," *School and Society,* Vol. 26: pp. 691-94, Nov. 26, 1927.

Wood, Ben D. *Measurement in Higher Education,* World Book Co., 1923.

Wood, Ben D. *New York Experiments with New Type Modern Language Tests,* The Macmillan Company, 1927.

CHAPTER XXII

Construction of Achievement Tests[1]

In constructing standardized tests one should follow generally recognized methods of procedure. The various steps in such a procedure are given here. Certain tests or scales may not include all of the steps listed. Most of these steps, however, are necessary in the scientific construction of standardized tests or scales of measurement.

1. To determine what is to be measured. This includes not only a definition of the outcomes of instruction or learning in any particular subject, but also a definition of the grade or grades and ages with which the measure is to be used.

2. To determine the criteria of success in the lines of work or the outcomes of instruction to be measured.

3. To select the items. In the light of the outcomes to be measured select a large number of items that would probably be associated with success in the desired outcomes.

4. To organize the items according to the technique best suited for testing, determining

(a) The type of question or exercise that is best suited to each item that is to be included.

(b) The organization of various types of questions for the purpose of testing. This last step will include some preliminary editing of the try-out test.

[1] It is the purpose of the authors to present in this chapter a simple statement and discussion of test development. Very little of the statistics of test development has been included. The reader trained in general statistics who desires to find its application to the special field of testing should consult such references as: T. L. Kelley, *Interpretation of Educational Measurements,* World Book Company, 1027; A. S. Otis, *Statistical Method in Educational Measurement,* World Book Company, 1925.

5. To give the experimental preliminary try-out of this large number of items in the form worked out under No. 4.

6. To study and analyze the results obtained from giving the try-out test.

(a) Analyze the statements and pupil responses.

(b) Study the results statistically to determine difficulty of items and per cent passing the successive grade levels.

7. To scale and weight the items of the test if this step is desirable. That is, to give more value to more difficult or important questions or items.

8. To break the items into equivalent and duplicate forms if these are desired.

9. To determine the equivalence of the duplicate forms.

10. To determine uniform methods of giving, timing, and scoring the test and of tabulating results.

11. To derive norms and standards.

12. To determine the reliability of the test as a measure of groups and of individuals.

13. To determine the validity of the test by correlating results obtained from its use with the results of other measures of the same type.

14. To edit the forms, giving due cognizance to cost in time, money, and ease of use.

15. To provide and edit a manual of directions.

16. To provide adequate keys.

To Determine What Is to Be Measured

The first step in the construction of an achievement test is to determine what is to be measured. This important question can be answered by a study of the objectives and aims of instruction in the school subject concerned. These aims in general determine legitimate outcomes of instruction and are therefore the things to be measured. The attainments expected in the various grades determine the limitations of the content

of a test for any grade or grades. In many school subjects not all of the aims or objectives of instruction can as yet be satisfactorily measured. In the development of a standard test in any school subject the objectives that are measured should be stated. The objectives that are not measured should also be stated.

The Criteria of Success

No other single phase of standard testing is so difficult as is the determination of the criteria of success. What constitutes success in any subject is difficult to know. In some few cases an analysis has been made. For example, in handwriting legibility the demands of social life have been determined. In most lines of work a score of 60 on the Ayres Handwriting Scale (Gettysburg Edition) is considered satisfactory legibility. It is satisfactory for the average needs of those vocations in which writing is much used. What are our criteria of success as applied to geography? What geography should an adult person know and understand? As the analysis of social needs is carried further, perhaps the answer to this question may be obtained. Even where it is possible to determine the criteria of success in any given line of work, say, for example, geography, it will still be necessary to determine the criteria of success in each grade or in each unit of organization of the school. What and how much geography should the average child at the end of the sixth grade know and understand? This question might be answered by determining how much the average child at the end of the sixth grade does know and understand. However, this is not an answer because it represents only present practice and that may not be success. There is often a difference between the criteria of success and the average achievement.

At the fourteenth annual meeting of the Music Supervisors' National Conference held in St. Joseph, Missouri, 1921, members of the Educational Council submitted a standard course of

study in music for graded schools. This course was formulated for the good or average school and was supposed to represent aims easily attainable by the end of the sixth grade. This course received the unanimous endorsement of the members of the Council after it had been "thoroughly analyzed and discussed." The purpose of the course was to direct the attention of educators to standards of attainments toward which the music work in schools generally should go.

First attainment. This attainment recommended by the Council for the sixth grade is that "every child shall have acquired the use of his singing voice and pleasure in song as a means of expression."

Second attainment. "Every child shall have acquired a repertoire of songs which may be carried into the home and social life, including 'America' and the 'Star-Spangled Banner.'"

Third attainment. "Every child shall have developed the aural power to know by sound that which he knows by sight and vice versa."

Fourth attainment. "Every child shall have the ability to sing at sight, using words, a unison song of hymn tune grade, and the easiest three-part songs; these to be in any key; to include any of the measures and rhythms in ordinary use; to contain any accidental signs and tones easily introduced; and in general to be of the grade of difficulty of folk songs such as the 'Minstrel Boy'; also knowledge of the major and minor keys and their signatures."

Fifth attainment. "Every child talented in musical performance shall have opportunity for its cultivation."

Sixth attainment. "The children shall have developed a love for the beautiful in music and taste in choosing their songs and the music to which they listen for the enjoyment and pleasure which only good music can give."

Seventh attainment. "The children shall have acquired the ability to appreciate the charm of design in songs sung; to give an account of the salient features of structure in a standard composition after a few hearings of it; to identify at least the three-part song form from hearing; and to recognize and give titles and composers of a reasonable number of standard vocal and instrumental compositions."

The average achievement in music at the end of the sixth grade is far below these criteria of success.

Forming a tonal image from a notational symbol is a basic skill in sight-singing. Test 1 of the Kwalwasser-Ruch Accomplishment Test measures one's knowledge of musical terms and symbols. This test consists of twenty-five items. The mean score of the children measured is 13. Evidently many of the most common symbols and terms are not understood by sixth-grade pupils.[2]

The criteria of success in any given line of endeavor may well be determined by (a) the social value of the achievement and (b) the degree of achievement that should be attained by the child working under the most favorable conditions of learning.

Selection of the Items

The selection of the original items of the test may be made in numerous ways. The most common ones are as follows:

1. Analysis of the content of the most commonly used textbooks in any given field.
2. Analysis of the best available courses of study.
3. Analysis of examination questions.
4. Consensus of opinion of competent judges, teachers, and others who have studied the material with which the test is concerned.
5. Combination of those methods.

Items that occur in one-half or more of a large number of the most commonly used textbooks in any subject can be said to be representative of the teaching practice. These items are very likely to be taught in any but the exceptional school. A test on those items is valid. At the same time the selection on this basis may not be fair to some more progressive schools or teachers whose work is in advance of common practice. It

[2] *Fifth Yearbook,* Department of Superintendence.

does not give a full measure of such work and in some ways may penalize it. A measure based on the items occurring in the most commonly used textbooks is a measure of the status quo and therefore does not suggest most progressive content or teaching method. Such a method of selecting items is practical especially for subjects that have recently come into the curriculum.

An example of this method of textbook analysis is given:

The selection of the items in the Ruch-Popenoe General Science Test has been based largely upon an analysis of the contents of existing textbooks in general science. Although there are more desirable criteria for the validation of test materials for most school subjects than the one here used, it is nevertheless true that, during the formative period of any new subject, teaching practice follows very closely the organization of the available textbooks. This very practical consideration is held to be sufficient justification for the construction of a test which embodies the best of the materials actually occurring in the books which must be used in general science classes.

The present form of the test is a revision and extension of the Range of Information Test in General Science published in 1919 by one of the authors. The original test consisted of three forms of fifty scientific terms each, which had been selected from a list of more than five hundred such terms obtained by analysis of twenty-three texts and manuals in general science. Of these five hundred items, about one hundred and eighty were common to half or more of the texts and for this reason were considered to be valid for test purposes. Some further eliminations, necessitated by reasons stated in the articles cited below, reduced the final number to one hundred and fifty. These were roughly equated and standardized as three forms of fifty items each.

In the present test, two of these forms have been standardized in more objective form. The ease of scoring has been greatly increased, and greater objectivity of scoring has been made possible.

The selection of the diagrams of Part II was carried out in a manner similar to the information item of Part I, except that

the ratings of a number of experienced teachers were secured as an additional guarantee of the desirability and validity of these materials. Some further eliminations were made in order to prevent undue overlapping of the fields covered by Parts I and II.[3]

An analysis of courses of study as a method of obtaining items to be included in a test is theoretically a good method. It is likely to represent practice of the latest type. However, the courses of study as they usually appear do not contain much detailed content. Such a method is better to determine aims and objectives than to give concrete material for the test builder. In the last analysis the test builder will find it necessary to go to the textbook and other material used in class to obtain the items to be included.

An analysis of examination questions has advantages that make its use desirable. Such questions should represent the judgment of teachers as to what the most important items are as they have taught them from the textbook in accordance with the course of study. One investigator[4] has found by the analysis of such questions in history that the questions are too factual and do not represent the best ideas of history teaching. However, the analysis of examination questions is both valid and usable if a wide enough range of questions is included in the analysis and if the teachers from whose questions the selections are made represent the highest type of teachers in the particular subjects.

The following quotation[5] is representative of the use of the analysis of examination questions method.

The selection and validation of the items in the Ruch-Cossmann Biology Test have been largely based on the analysis

[3] Manual of Directions for *Ruch-Popenoe General Science Test* (World Book Company), 1923, pp. 1-2.

[4] W. J. Osborne, *Are We Making Good at History Teaching?* Public School Publishing Company, 1925.

[5] Manual of Directions for Ruch-Cossmann Biology Test (World Book Company), 1924, p. 2.

and condensation of two thousand final examination questions obtained from a collection of final examinations used in all parts of the United States over a period of several years.

The selection of items by the consensus of opinion of competent judges is in many respects represented by any of the methods just mentioned. The textbook, the course of study, the examination questions, are on the whole the opinions of competent people arrived at by study and investigation. The advantages of obtaining the pooled judgments of competent teachers or workers in any field or subject are: (a) such a method may be used to add more competent judgments to those that are expressed in textbooks and courses of study; (b) such a method gives proper weight to the ideas of more progressive teachers and workers in any given field or subject.

Selecting items by the method of judgment or consensus of opinion may be done in two ways. The competent teachers and workers in any given field may make an extensive list of items from which may be selected those most commonly listed or these people may serve as judges of an extensive list of items selected by the other methods indicated. The second method is likely to be the most valuable one in practice. The following method can be used for determining the general merit of the items of a large original list of items. Place each item on a separate card; have each judge study each item carefully. Then upon the basis of the merit of the item as a part of a test to measure achievement in any given subject the judge should determine if the item is either (1) Entirely satisfactory, or (2) Satisfactory, or (3) Poor and should not be used. If the item is entirely satisfactory, the judge should turn the card over and mark a figure 1 on the back. If it is only fairly satisfactory, a 2 should be marked on the back of the card. If it is poor and should not be used, a 3 should be marked on the back of the card. The judgment should be made before the card is turned over. In this manner a large number of judgments can be secured.

It is true that the value of this method of pooled judgments or consensus of opinion will depend upon the competence of those making the judgments. However, carefully selected teachers and workers in any field may well represent as good a method for validation of items as any that is attainable in most fields at present.

In some subjects of the public schools scientific studies have been made that should be used as a final check upon any method of selecting items for a standard test. If a test of word knowledge is desired for grades 4 or above, it would hardly be advisable to select the words to make up the test without using, as a final check, at least the words found in *The Thorndike Teacher's Word Book*.[6] In determining the places to be located in geography, the scientific criteria found in *Objective Studies in Map Location*[7] should be used as a final check. When other fields of school work have been as fully investigated as the two mentioned, it will be possible to use the results of these investigations as final checks on any method of selecting items for a standard test.

A combination of the various methods suggested for the selection of the original items of the test will, of course, be found to be most accurate and satisfactory.

Organization of the Items for the Try-out Test

After the items of the test have been selected, the next step is to turn these items into the form of a test. As was explained in the previous chapter, this requires a study of the objective type questions to determine what type of question is best suited to each item.

The number of items should include at least twice the number that will later be included in the final forms of the test.

[6] Edward L. Thorndike, *The Teacher's Word Book*, Teachers College, Columbia University.

[7] H. Rugg and J. Hockett, *Objective Studies in Map Location*, Lincoln School of Teachers College, Columbia University.

This will allow for elimination of those items that are found to be ambiguous in statement, too easy, too hard, etc.

The questions on the try-out form of the test should be so arranged that those of the same type will come together. The multiple response-single choice questions should come together, the true-false questions together, and in the same manner the other types of questions.

The same general rules stated for the organization of the new-type or objective examination as set forth in the previous chapter apply to the try-out form of the test.

The Preliminary Try-out of the Test

After the items have been properly turned into objective type questions and organized according to the rules set forth for new-type or objective tests and printed or mimeographed, the next step is the giving of the try-out form. The test should be given to a sufficient number of pupils of the desired type. As an example, if a geography test is desired for use in the elementary school, it would be necessary that pupils of the 3rd, 4th, 5th, 6th grades of normal and typical schools be used. If it is not possible to use all the pupils of the grades of various schools, the number can be reduced by sampling in either of two ways. (a) Have the teacher rank the pupils according to her judgment in the order of their accomplishment in geography. Then, beginning with the median pupil in each group, work both ways until the number desired is selected. (b) Select every other pupil alphabetically. The number of pupils from each grade should not be less than one hundred, and it would be better in each case to have many more than that number.

In the preliminary try-out the test should be given in as careful and exact manner as a standard test with the exception that time should be allowed sufficient for each pupil to attempt all the items. If there is need for conserving time, the test

may be stopped when a few very slow pupils have not yet completed the work.

Analysis of the Results of the Try-out Test

After the test has been scored, the results obtained in the try-out of the preliminary form should be studied in two ways: (a) By an analytical and critical study of the pupil responses; (b) by a statistical study. The first method requires that the response to each question answered incorrectly be studied to determine if the pupil answered the question incorrectly because he did not understand the question or because he did not know the correct answer. This is, after all, a study of the ambiguity of the various questions and directions from the point of view of the pupil's interpretation of them. It is, further, a study of the value of the various questions as measures of the outcomes of instruction. The questions should be studied also to determine if there are responses, other than those listed in the key, that may be satisfactory. The responses may also be studied to collect those that are definitely wrong to be used as examples of wrong answers. A list of possible correct responses and a list of examples of incorrect responses are needed to use with those tests that are not completely objective. Such a study of pupil responses will usually lead to the discarding of many questions found to be ambiguous in statement as indicated by pupil responses, or lacking in objectivity, or inadequate as a part of the measure of a desired outcome.

The statistical study of the test should be made (a) to determine per cent passing each question for the various grades (or age levels) in which it was tried out, (b) as a preliminary check on reliability. Both of these steps are for the purpose of determining the number and type of items to be retained in the test.

TABLE 59.—THE PER CENT OF PUPILS PASSING EACH ITEM OF A TRY-OUT TEST AT SUCCESSIVE GRADE LEVELS

Item No.	*Per Cent Passing in Grade*					
	III	*IV*	*V*	*VI*	*VII*	*VIII*
1	10	13	18	22	27	33
2	4	18	33	54	76	98
3	0	0	5	8	12	15
4	0	14	30	58	36	19
5	24	59	93	100	100	100
6	0	0	0	6	39	72

Study of a table like Table 59, showing all the items of the try-out test that might be considered for the purpose of inclusion in the final form, reveals the following points:

(a) Item No. 1 is passed by 10 per cent of the pupils of the third grade and by a slightly larger per cent of each succeeding grade. The question is somewhat difficult for all of the grades tested.

(b) Item No. 2 is passed by a small per cent of the pupils of the third grade, and by an increasing per cent of each successive grade; and the differences between grades are greater than found in item 1. The question is rather difficult for the third grade and rather easy for the eighth grade. From the point of view of per cent passing successive grades, this question would be quite satisfactory for a test covering these six grades.

(c) Item No. 3 is too difficult for grades III and IV.

(d) Item No. 4 is unsatisfactory because it does not give a consistent rise in per cent passing in successive grades. In this item an average sixth grade pupil would do better than an average eighth grade pupil. This will tend to make the test unreliable or unnecessary.

(e) Items Nos. 5 and 6 represent a rapid rise in per cent passing. This means that these items tend to distinguish sharply between the pupils of the various grades. A test made up of this type of question would not have as much overlapping

in the distribution of the scores of two successive grades as a test made up of items like No. 1 and No. 2. If the test is to be used in a narrow range of grades, i.e., III-IV or VII-VIII, such items as 5 and 6 should be chosen. Item 2 would be as satisfactory for the range of grades from III-VIII.

By a study of Table 59 the relative difficulty of each item as it relates to the other items can be ascertained. The items on the test should come in order of increasing difficulty. The item of the test passed by the largest per cent is supposed to be the easiest. It would, then, be placed first. The item passed by the smallest per cent would come last. The per cent passing each item would decrease from the first to the last items on the test. By this method the items as they appear on the final form will be graduated as to difficulty.

Another rough method of determining the value of the items of the test as a measure of achievement is to select the twenty poorest and the twenty best pupils in a group of one hundred as designated by the teachers and find their responses on certain items. An example is given in Table 60.

TABLE 60.—DETERMINING THE GOODNESS OF TEST ITEMS

Item No.	*No. of Correct Responses of 20 Poorest Pupils*	*No. of Correct Responses of 20 Best Pupils*	*Difference*
1	12	14	+2
2	9	17	+8
3	11	12	+1
4	9	9	0
5	8	4	−4
6	6	18	+12

Wherever the difference is + and much in favor of the best pupils, the item is said to have that degree of "goodness." Items 2 and 6 are outstandingly good; items 4 and 5 are not valuable and therefore "bad." The measure of "goodness" of any particular item is the number of correct responses of the

best pupils minus the number of correct responses of the poorest pupils.

A further analysis of the results of the test should be made to determine the approximate reliability to be attained with the satisfactory test items.[8] Tests of low reliability are of less value for accurate measurements than are tests of high reliability. It is therefore well to set a minimum reliability to be attained. One of the factors of high reliability is the length of the test. Other things being equal, the longer the test the higher the reliability. Therefore, a preliminary measure of reliability should be taken to determine the number of items to be used in a single form of a test. One way to test reliability from the try-out form is to find the total score made on every other item that has been found to be valid for use in the test, and correlate that score with the total score made by the same pupil on the remaining valid items. This has the effect of dividing the items into two tests, each made up of one-half of the items, these having been selected by chance. Take an example of 120 valid items. The plan would be to find out the score the pupil made on items 1-3-5-7-9, etc., and then on items 2-4-6-8-10-12, etc. The next step would be to find the correlation between these two scores (evens *vs.* odds) with a large number of properly selected pupils. If this method were followed, the coefficient of correlation thus found could be used to estimate the reliability of the test if the whole number of items were used as one test. To do this the Spearman's (sometimes called Brown's) prophecy formula, $r_x = \frac{Nr}{1 + (N-1)r}$, should be used. Let it be assumed that the correlation of the scores of pupils on the even and odd items was .82. What would be the estimated reliability if all the items were taken as one test? In the formula just given $r_x =$ the estimated coefficient of reliability of the whole test, if all the items of the present form are used.

[8] See discussion of reliability on p. 535.

$N = 2$. In this case the number of items is doubled. That is, the even and odd are both included.

$r = .82$. This is the coefficient of correlation between the even and odd items of the whole number of items.

Substituting in the formula

$$r_x = \frac{(2)\ (.82)}{1 + (2 - 1)\ (.82)} = \frac{1.64}{1 + .82} = \frac{1.64}{1.82} = .901 \text{ or } .90.$$

This is the estimated reliability of the test as a whole. Thus it may further be noted that doubling the length of the test increased the reliability of the test from .82 to .90 since .82 might be called the reliability of either the evens or the odds or one-half of the whole 120 items.

If a reliability coefficient of .95 is desired it is possible to determine how much longer than the present length the test needs to be to get the desired reliability coefficient. This is, of course, assuming that the same quality and type of items would be used.

Substituting in the same formula and solving for N gives

$$.95 = \frac{(N)\ (.90)}{1 + (N - 1)\ .90}$$

The formula can be rearranged so as to give N when the desired reliability is known and the observed reliability is given.

$$N = \frac{.95\ (1 - .90)}{.90\ (1 - .95)} = 2.1+$$

.95 is the desired reliability.

.90 is the reliability on the whole test of 120 items.

N is to be solved for. This is the number by which one should multiply the present number of items (120) to obtain the desired reliability of .95.

The test would have to be 2.1 times as long as 120 items, i.e., have in it 252 items, in order to have a reliability of .95.

In this manner the length of the test or the number of test items needed, of the same quality, can be ascertained. In the example given, the solution for increasing reliability might not be found in the lengthening of the test alone since the number of items necessary to obtain a reliability coefficient of .95 makes the test too long to use easily under some conditions.

Scaling and Weighting the Test

In general it has been found impractical to weight items on an ordinary test or product scale. While it would seem theoretically sound to give more value or weight to the difficult items, there are two reasons for not doing so. (1) In a test with a large number of well-selected items it is much simpler and at least approximately as accurate to use raw scores. (2) Increase in difficulty due to a low percentage of successes does not necessarily show the intrinsic difficulty of any question. The difficulty may arise from unimportance or rarity of the question. If the latter is true, weighting would give overimportance to items of little significance.

In quality scales like the handwriting, drawing, composition, and sewing scales, the general practice is to weight the various samples.

Equivalent Forms

If the items selected from the analysis of the results of the try-out have been put in order of increasing difficulty, and if it has been determined that one hundred items are sufficient to give the reliability desired, the next step is to divide the items into equivalent forms. Let it be assumed that there are three hundred valid items available and that three equivalent forms of the test are to be prepared. To make them equivalent, the following method should be used (Table 61).

If the items increase in difficulty and Form *A* should consist of items 1-4-7-10-13-16-19, Form *B* of 2-5-8-11-14-17-20, and

Form *C* of 3-6-9-12-15-18-21, the three forms would not be equivalent in difficulty. In this arrangement item 1 of Form *A* is easier than 2 of Form *B* and item 4 of Form *A* is easier than 5 of Form *B*, etc. From this it can be seen that each item on Form *A* would be easier than the corresponding item on Form *B* and each item on Form *B* easier than the corresponding item on Form *C*. Consequently the three forms are *not* equivalent. If, however, the method of arrangement given in Table 61 is followed, the forms are equivalent. Item 1 of Form

TABLE 61.—AN ILLUSTRATION OF THE METHODS OF EQUATING FORMS OF A TEST

FORM A *Item No.*	FORM B *Item No.*	FORM C *Item No.*
1	2	3
6	5	4
7	8	9
12	11	10
13	14	15
18	17	16
19	20	21

A is easier than item 2 of Form *B*, but item 5 of Form *B* is easier than item 6 on Form *A*. Likewise, item 3 on Form *C* is more difficult than either item 1 on Form *A* or item 2 on Form *B;* but item 4 on Form *C* is easier than either item 6 on Form *A* or item 5 on Form *B*. In this manner the relative difficulty of items is balanced. The medians, means, and standard deviations of these three forms would then be very nearly the same.

A check must constantly be kept on the items that go into the various forms in order to prevent biased measurement. If the test covers the whole field of elementary geography, it would be possible, although not likely, that all the items concerning any one continent would be in one form only. This is a type of bias to be avoided by checking the content of the three forms.

Regiving the Test

The test, in printed form, should be given again (a) to determine the equivalence of duplicate forms, (b) to establish final methods of timing, scoring, and tabulating the results of the test, (c) to derive norms and standards, (d) to determine the final reliability of the test as a measure of groups and individuals, (e) to further validate it.

Equivalence of Forms

If the various forms of the test are now given to the same children in the various grades in which they are to be used and the results compared, it will be possible to determine the equivalence of the forms. The means and standard deviations of the forms should not be different from each other more than one-half a grade if they are to be called equivalent. If they are not equivalent, they can be corrected so as to be considered equivalent by adding a constant to one form. For example, if the score on one form ran five points lower than that of another, these two forms could be made equivalent by always adding five to the score of the test whose scores ran lower each time the test was given.

Time Limit

The final time limit in which the test is to be taken is dependent upon the type of test. In general there are two types of tests. One is what is called a *rate* test, the other is usually called a *power* test. A rate or speed test should be accurately timed. No pupil is supposed to finish in the allotted time. The items on which the pupil is tested are supposed to be of equal difficulty. Where it is desired to measure rate as such, the test should be accurately timed. The rate of reading, the rate of handwriting, and the rate of working arithmetic prob-

lems should be so determined. The score obtained on such a test is a measure of the amount done in a given unit of time. The power test is supposed to measure not how fast the pupil can work, but how difficult a task he can do. The time element is not important. Since in most subjects it is desirable to measure the degree of mastery rather than the speed of work, it seems rather desirable in most tests to allow a generous amount of time. Because a few pupils are likely to be very slow, it is, however, well to set a time limit. What this time limit should be, can be determined rather fairly by a try-out method. Assume (from best judgment) that the test will probably take 20 minutes. Have the pupils instructed that when the examiner says "mark" they are to put a circle around the item they are working on and go ahead working. The examiner can say "Mark" at 15 minutes, again at 18 minutes, again at 20 minutes, again at 22 minutes, and at 25 minutes say "Stop." The time needed for approximately 90 per cent to finish can be used as the time for the test. Still another method is to tell each pupil to hold up his hand when he has finished. When approximately 90 per cent have finished, say "Stop." By keeping a record of the time elapsed from the time the pupils began work, the time can be set for the test.

Still another method, not so easily used but more accurate, is for the examiner to write on the board the minutes as they elapse from the time of beginning the test. The pupil is instructed to look at the board as soon as he has finished working and to copy from the board the last number indicating the number of minutes elapsed. An analysis of the time thus required to finish the test will give a basis for a suitable standard.

Norms

There are four kinds of norms usually used. They are grade norms, age norms, percentiles, T-scores. To obtain *grade norms* it is necessary to give the test to a large number

of pupils that are representative of a grade. The grade norm is widely used in the elementary schools and will probably continue to be. Grade norms are not useful in high school subjects. Because the grade concept and practice are common and because the grade norms are easily obtained, it is suggested that this norm be the first one to be determined in any test of achievement in elementary school subjects.

The norm of any grade is usually the average score of those tested as representative of that particular grade. The grade norms should be given by units of the grade whenever possible. A score of 17 in arithmetic may represent the norm for the end of 4A. Such a device as the *B* score inaugurated by McCall may be useful as a grade norm practice. The *B* score is obtained by dividing the increase in grade-to-grade average scores into tenths. Then any score can always be translated into terms of grade norms. For example, a score of 17 in an arithmetic test may be translated into a 4.5 grade norm. As tests become more reliable and differentiate better between grades, the *B* score will become more valuable. It is most valuable now when used with such composite tests as the Stanford Achievement test and the Indiana Composite Achievement test.

Grade norms have certain inaccuracies that limit their usefulness. Grades do not mean the same in all places. In some parts of our country the elementary schools and high schools consist of 7 and 4 years respectively, in others 8 and 4, in still others 6 and then 3 of junior high school and 3 of senior high school. Manifestly grade norms for grades above the sixth grade have a doubtful meaning. Another type of inaccuracy that makes the grade norm somewhat less useful is that the practice of promotion in some schools differs markedly from the practice in others. If pupils in one school are accelerated and in another acceleration is not practiced, a grade norm will hardly fit both cases.

Age norms are obtained by getting the average achievement of all the pupils of a given chronological age regardless of

grade. If the average achievements of all 10-year-olds, 11-year-olds, and 12-year-olds are obtained on any test, then by interpolation it is possible to obtain the intervening age scores. The average age of 10-year-olds is 10 years 6 months, of 11-year-olds 11 years 6 months, and of 12-year-olds 12 years 6 months. Suppose the following age norms were obtained on a reading test for these groups.

Age	10-6	11-6	12-6
Score	40	52	64

It would be easy to interpolate and find as follows:

Age....	10-6	10-7	10-8	10-9	10-10	10-11	11-0	11-1	11-2	11-3	11-4	11-5	11-6
	40	41	42	43	44	45	46	47	48	49	50	51	52

In this manner some tests now give scores in terms of subject age or, in case of composite tests, educational age by years and months.

Such age norms properly arrived at allow a comparison with mental age and chronological age not possible with grade norms. However, such norms are not easily obtained and on the whole are not easily understood. Age norms have not been found successful in use with high school tests because a large number of older pupils drop out of school before they reach high school.

Another type of achievement norm that is being used to some degree is the *mental age achievement norm.* This is found by taking the average score on an achievement test of the pupils of a certain mental age. The average achievement score in reading of all the pupils with a mental age of 11 (from 11-0 to 11-11) would be an example of the mental age achievement norm. This has some value as a means of interpreting achievement in terms of ability.

From many points of view, the *percentile* is the most satisfactory type of norm, especially with tests used above the sixth grade. Percentiles are easily understood and not difficult to cal-

culate if the procedure is understood. The following example will give an understanding of a simple method to be used in turning raw scores into percentiles.

TABLE 62.—A METHOD OF DETERMINING PERCENTILE EQUIVALENTS FROM RAW SCORES

Scores	*Frequency*	*Subtotals*	*Operation*	*Percentile*
69	1	50	50 ÷ 50 × 100 =	100
66	1	49	49 ÷ 50 × 100 =	98
64	1	48	48 ÷ 50 × 100 =	96
61	1	47	47 ÷ 50 × 100 =	94
59	1	46	46 ÷ 50 × 100 =	92
58	1	45	45 ÷ 50 × 100 =	90
57	1	44		
56	1	43		
55	1	42		
54	1	41		
53	3	40	40 ÷ 50 × 100 =	80
52	2	37		
50	3	35	35 ÷ 50 × 100 =	70
49	2	32		
48	4	30	30 ÷ 50 × 100 =	60
46	1	26		
45	2	25	25 ÷ 50 × 100 =	50 = median
44	1	23		
43	2	22		
42	3	20	20 ÷ 50 × 100 =	40
41	2	17		
40	1	15	15 ÷ 50 × 100 =	30
39	2	14		
36	2	12		
35	2	10	10 ÷ 50 × 100 =	20
34	1	8		
33	1	7		
32	1	6		
31	1	5	5 ÷ 50 × 100 =	10
28	1	4	4 ÷ 50 × 100 =	8
26	1	3	3 ÷ 50 × 100 =	6
23	1	2	2 ÷ 50 × 100 =	4
18	1	1	1 ÷ 50 × 100 =	2

The subtotals given in Table 62 are obtained by adding the frequencies, beginning at the bottom, to any point. For example, adding all the frequencies up to and including the score of 35, the subtotal is 10. This means that ten cases (or pupils) of the 50 that took the test made a score of 35 or less.

Adding the column of frequencies up to and including the cases (or pupils) that made a score of 48, the subtotal is 30. This means that 30 of the 50 pupils who took the test made a score of 48 or less.

To obtain the percentile (the per cent that made a given score or less) it is necessary to divide the subtotal by the total number of cases and then to multiply the result by 100. For example, if the frequencies are added up to and including a score of 31 the subtotal is 5. Divide this number by the total number of cases, which gives .1. Then to get this into per cent multiply by 100. This gives a percentile of 10, which means that 90 per cent of the cases in this group make scores above 31 while 10 per cent make scores of 31 or below. Likewise, by adding the frequencies from the bottom to include the cases up to and including the score of 59, the subtotal is 46; this divided by 50, the total number of cases, gives .92; this multiplied by 100 gives 92 found in the table under percentile. This interpreted means that 92 per cent of the group in question made a score of 59 or less while 8 per cent made a score of above 59.

This simple method of finding the percentile gives results that interpret each score specifically in terms of the grade or age of the special group that has been selected as representative. A method of graphing the results of the test on a universal percentile graph has been worked out in some detail and published.[9]

The *T-score* is a score in terms of standard deviation of the distribution of scores. Standard deviation is a measure of variability of the scores of a distribution. That is, it is a measure of how much the scores vary from the mean score. This unit of variability is called a standard deviation or sigma. The T-score is, then, a score in terms of variation from the average. It is necessary to calculate the mean and the standard

[9] Otis, Arthur S. World Book Company.

deviation before the T-scores can be derived. The following formula is used in deriving the T-score.

$$T = 50 + \frac{10\,(X - M)}{\sigma \text{ (Standard deviation)}}$$

T means T-score.
X means any raw point score of a test.
M means the average or mean of the distribution.
σ or Standard deviation is the square root of the sum of the squares of the deviations from the mean of the distribution.

The 10 is used merely to eliminate fractions and 50 is used to make the average T-score 50; the other T-scores will be between 0 and 100 or above and below 50. The following example will illustrate how the raw point scores can be transferred into equivalent T-scores (Table 63).

TABLE 63.—THE CHANGING OF RAW POINT SCORES INTO T-SCORES

Raw Score (X)	*Deviations of Raw Scores from the mean* ($X-M$)	*Integral Values* $\frac{10\,(X-M)}{\sigma}$	*T-scores* $50+\frac{10\,(X-M)}{\sigma}$
10	4	13	63
9	3	10	10
8	2	7	57
7	1	3	53
6	0	0	50
5	—1	— 3	47
2	—4	—13	37
1	—5	—17	33

Mean = 6.
Standard deviation (estimated) 3. Figures in nearest whole numbers.

A T-score can in this manner be determined for any group of individuals. However, McCall suggests that for certain valid reasons a standard group should be used. This group he suggests should be unselected 12-year-old pupils.

Reliability

The reliability is a measure of the degree to which a test measures consistently whatever it does measure. If two or more forms of a test have been developed, it is advisable to determine the reliability of these forms in the following manner. (a) Test a typical group of more than 100 pupils with both tests on succeeding days at the same time of day and under the same conditions. (This part of the work probably will have been done in testing the equivalence of the forms if more than one form has been developed.) (b) Score the test according to the directions for scoring. This will give a score on each of the two equivalent forms for each pupil tested. (c) Correlate the scores made by the same pupils on the two different forms. The correlation obtained is a measure of the reliability if the forms are equivalent.

The reliability coefficient is a general measure that tends to show the trend of behavior of the test when used with a large number of pupils. As the coefficient increases, the indication is that such tests will tend to give more nearly the same result if used again or that one equivalent form will give more nearly the same result as another.

The reliability coefficient does not tell the error of the test on an individual pupil's test score. The amount of error in an individual's score is usually stated as the probable error (P.E.) of the raw score. The amount of this error can be obtained from the following formula:

$$\text{P.E. (raw score}^{10}) = .6745\ \sigma_1 \sqrt{1 - r_{12}}$$

In this formula r_{12} is the correlation between the two forms of the test. σ_1 is usually the average standard deviation of the two forms of the test. Let us assume that the two forms of the test have been correlated and that they give a reliability coefficient of .91. Let us further assume that the standard

[10] An individual score made on the test.

deviation of one form is 12 and of the other, 10. By substituting in the formula, we should have the following:

$$\text{P.E. (raw score)} = .6745 \frac{10+12}{2} \sqrt{1-.91}$$
$$\text{P.E.} = 7.4195 \sqrt{.09}$$
$$\text{P.E.} = 2.22$$

Let us assume that a pupil made a score of 50 on a test. The meaning of this score in the light of the probable error given may be interpreted as follows: The chances are 1:1 that the true score of this pupil lies between 47.88 and 52.22, that is, within a range of 2.22 in either direction from the raw score. The chances are 4.6:1 that the true score lies between 50 minus 2 (2.22) and 50 plus 2 (2.22); that is, between 45.56 and 54.44.

Let us further suppose that the fourth grade pupil has made a score of 40 on a test which has the following norms:

	III	*IV*	*V*	*VI*
Form *A*.....	30	40	50	60
Form *B*.....	30	41	52	60

The score of 40 made by the fourth grade pupil on Form *A* of the test can now be interpreted more specifically. The average difference between grades given in the table is approximately 10. The P.E. of the raw score is 2.22. The ratio of the probable error of the pupil's raw score to the difference between the two grades is $\frac{2.22}{10}$. The chances are very strong that the error of measurement would not be enough to place this fourth grade child in any other grade norm. If the reliability coefficient in the preceding formula had been .64, the probable error of measurement would have been 4.45. The chances are 4.6:1 that the true score would be between 41.1 and 58.9. Therefore it can be seen that with as low a coefficient of reliability as .64 the individual pupil might be placed approxi-

mately one grade too high or one grade too low by the error in a score on the test. In this manner the coefficient of reliability gives some measure of how accurately a pupil may be placed by his individual score. Two important measures are related in finding this probable error of the raw score. They are the standard deviation of the distribution and the reliability coefficient. If the standard deviation increases, the probable error of measurement will increase. As the coefficient of reliability decreases, the probable error of the score increases.

Validity

The validity of a test may be checked by seeing how closely its measures are like other standard measures of the same thing. To determine the validity of a narrative reading test, it would be well to find out how closely the measure of ability to read narrative material as measured by the new test correlates with the measure of the same thing by other tests whose function in measurement is the same. The best single measure or check by statistical method on the validity of a test is to correlate the score made by a pupil on the new test with the composite score made by the same pupil on several other like standardized tests. The test can also be validated statistically by correlating the scores made on the test with the school marks given by teachers.

Editing the Test

This important step has been neglected in many cases. Two things may be accomplished by editing the test. One is that the cost of the test, both in money and in time, may be reduced. Another is that the test can be arranged, organized, and printed in accordance with the manner in which it is to be used. If the test is to be scored by the pupil instead of the teacher, it should be edited accordingly.

A great deal of time can be saved by making the test so

that it can be rapidly scored. A multiple response type of question in which the pupil writes, on the line to the right of the page, the number representing his response can be scored much more rapidly than if the response is just underlined. This fact should be made use of, in editing a test, so that the tests may be more rapidly scored. Keys should be so edited that they may be easily and handily used. Other things being equal, the cheapest test is the best one. It may be possible by proper editing to reduce the cost of testing to a marked degree.

The editing of the test should include careful work on the method of giving the test. The directions to the pupil should be clear and specific. These directions should be printed on the test. The directions to the examiner should not be on the test but should be in the Manual of Directions that accompanies the test. If the test is made up of various parts which have questions of different types, each part should be preceded by an example showing the pupil how to mark or indicate his answer. If the test is somewhat involved, the test should have in addition to the instructions "warming-up exercises."

Manual of Directions

The manual of directions or examiner's guide should be well worked out and carefully edited, and should include: (1) Name of test and authors, (2) Purpose of test, (3) The basis for the construction of the test, (4) Description of the test, (5) Standardization of the test, (6) Methods of validation, (7) Reliability of the test, as a measure of groups and of individuals, (8) Directions for giving the test, (9) Directions for scoring the tests, (10) How to interpret the results in terms of as many type norms as would be usable, (11) Uses of the results in various phases of school work, such as classification, diagnosis, administration, supervision.

Keys

Accurate, carefully edited, convenient scoring keys should be provided with the test. Directions for scoring should be printed on the key whenever possible. If more than one correct response is to be allowed as an answer to any item, the correct responses and examples of incorrect responses should be given.

BIBLIOGRAPHY

Bixler, H. H. *Check Lists for Educational Research.* Bureau of Publications, Teachers College, Columbia University, 1928, pp. 25-38.

Douglas, H. R. and Spencer, P. L. "Is It Necessary to Weigh Exercises in Standard Tests?" *Journal of Educational Psychology,* Vol. 14: pp. 109-12, February, 1923.

Greene, Harry A. "A Comparison of Two Methods of Evaluating Test Material," *School and Society,* Vol. 11: pp. 777-80, June 26, 1920.

Kelley, T. L. "A Simplified Method of Using Scaled Data for Purposes of Testing," *School and Society,* Vol. 4: pp. 34-37, 71-75, July 1 and 8, 1916.

McCall, W. A. *How to Measure in Education,* Chap. 4: pp. 122 ff.; Chap. 7 and 8, pp. 195 ff., The Macmillan Company, 1922.

Odell, C. W. "An Omnibus Achievement Test for the Upper Elementary Grades," *Elementary School Journal,* Vol 23: pp. 353-58, January, 1923.

Pressey, Luella W. "Scale of Attainment No. 1. An Examination of Achievement in the Second Grade," *Journal of Educational Research,* Vol. 2: pp. 572-82, September, 1920.

Ruch, G. M. and Stoddard, George D. *Tests and Measurements in High School Instruction,* World Book Co., 1927, pp. 301-75.

Sandiford, P. "The Standardization of Tests and Scales," *Journal of Educational Research,* Vol. 7: pp. 14-27, January, 1923.

Thurstone, L. L. "Method of Scaling Psychological and Educational Tests," *Journal of Educational Research,* Vol. 16: pp. 433-51, October, 1925.

Wood, Ben D. "Studies of Achievement Tests," *Journal of Educational Psychology,* Vol. 17: pp. 1-23, 125-39, 263-69, January, February, April, 1926.

MISCELLANEOUS

Laycock, S. R. "The Laycock Test of Biblical Information," *Journal of Educational Psychology,* Vol. 16: pp. 329-34, May, 1925.

MacCaughey, Vaughn. "The California State Survey of Mental and Educational Tests," *School and Society,* Vol. 22: pp. 210-11, August 15, 1925.

Melcher, George. "Suggestions for Experimental Work," *National Society for the Study of Education 17th Yearbook,* Part II, Chap. XI: pp. 139-51, 1918.

Patterson, M. V. W. "Effect of Summer Vacation on Children's Mental Ability and on Their Retention of Arithmetic and Reading," *Education,* Vol. 46: pp. 222-28, December, 1925.

Pusey, E. D. "Use of Educational Tests in Seven Grade School Systems," *High School Quarterly,* Vol. 12: pp. 237-38, July, 1924.

Sampson, Charles H. "Psychological Test as the Teacher's Judgment," *Educational Review,* Vol. 65: pp. 16-17, January, 1923.

Smith, C. M. "Study to Determine the Effectiveness of Kindergarten Training," *Elementary School Journal,* Vol. 25: pp. 451-66, February, 1925.

Symonds, Percival M. "An Index of Studiousness," *The Pedagogical Seminary and Journal of Genetic Psychology,* Vol. 32: pp. 257-65, June, 1925.

The following are the chief centers for the distribution of tests:

The Bureau of Educational Research and Service, University of Iowa, Iowa City.

Bureau of Public Personnel Administration, Mills Bldg., Pennsylvania Ave. and Seventeenth St., Washington, D. C.

Bureau of Publications, Teachers College, New York City.

Catholic Education Press, 1326 Quincy St., N.E., Brookland Station, Washington, D. C.

Harvard University Press, Cambridge, Mass.

Public School Publishing Co., Bloomington, Illinois.

Research Publishing Co., Washington, D. C.

Research Service Co., 4259 South Van Buren Place, Los Angeles, California.

C. H. Stoelting Company, 424 North Homan Ave., Chicago, Ill.

University Publishing Co., Chicago, Illinois.

World Book Company, Yonkers-on-Hudson, N. Y.

The following reference works contain elaborate lists of tests or bibliographies on tests or both:

Odell, Charles W. "Educational Tests for Use in Elementary Schools, Second Revision." University of Illinois Bulletin, Vol. XXIV, *Educational Research Circular No. 49.*

Odell, Charles W. "Educational Tests for Use in High Schools, Second Revision." University of Illinois Bulletin, Vol. XXIV, *Educational Research Circular No. 48.*

Smith, Henry Lester and Wright, Wendell W. *Bibliography of Educational Measurements* (Second Revision). The Bureau of Co-operative Research, Indiana University, November, 1927.

INDEX

Abbott-Trabue Exercises in Judging Poetry, 160; example of, 161
Achievement: educational chart, 18
Achievement Tests: age norms—method of obtaining—value of, 530, 531; analysis of the results of the try-out test, 521; analytical and critical study, 521; construction of, 511; criteria of success, 513, 514; determine reliability, 524; determine what is to be measured, 512; editing the test, 537; equivalence of forms, 528; example of method of selecting test items, 516; goodness of test items, 523; grade norms—how determined—limitations, 529, 530; how to obtain equivalent forms, 526; interpretation of errors of measurement, 536; keys, 539; length of test in relation to reliability, 525; manual of directions, 538; mental age achievement norm—how determined, 531; method of determining percentile, 532; norms—kinds, 529; per cent of pupils passing items on try-out test, 521, 522; power test, 528; preliminary try-out of the test, 520; probable error of the raw score, 535; purpose of re-giving the test, 528; rate test, 528; raw point scores transferred into T-scores, 533, 534; reliability, 521, 535; scaling and weighting the test, 526; selecting items, 515; selection of items by consensus of opinion, 518; statistical study, 521; steps in constructing, 511; time limit, 528; T-scores, 533; universal percentile graph, 533; use of analysis of examination questions, 517; validity—checked by, 537; value of percentile, 531
Algebra: ability with equations, 111; diagnosis in, 119; exercises and practice tests in algebra, 436; measurement deficient in, 111; objectives in, 110; testing should include, 112
Algebra tests (*see also name of test*), 112-114; 114-117; 118, 436
American Council Alpha French Test: administration of, 354; description and derivation, 353; excerpt from, 355, 356; purposes and uses of, 353; silent reading test, 354; vocabulary test, 353; written composition, 354
American History: excerpt from Van Wagenen Test, 244-246; intercorrelation of tests, 242; Pressey-Richards Tests, 250, 251; scales for rating pupils' answers to nine types of thought questions, 253; Van Wagenen Scales, 243
Arithmetic: abilities and skills in, 69; analysis of abilities in, 69, 70; analysis of work in, 13; computation test, 73; diagnostic tests in, 71, 82; errors in, 101; general survey tests in, 72; lists of tests in, 103; measurement of achievement in, 70, 71; objectives of, 68; problem solving in, 70; reliability coefficients of, 74; quotient of, 81; reasoning problem tests in, 71, 74; remedial measures in, 102; scope of complete testing program (Compass Diagnostic Tests), 82-86; Stanford examination in, 73; tests of general type in, 71; types of errors in reasoning problems, 100
Arithmetic tests (*see also name of test*), 103-106
Arithmetic Work-Book: criticisms of, 435; diagnosis of pupil diffi-

culties with, 434; good characteristics of, 435; progress chart used with, 434; purposes of, 430; use of, 430; work unit of, 431, 433
Art: 296, 297, lists of tests in, 299, 300
Ashbaugh, E. J.: types of persisting disabilities, 158
Athletic Badge Tests, 419
Ayres, L. P.: Laggards in Our Schools, 2
Ayres Handwriting Scale: description of scale, 186; diagnostic chart of, 188; sample of, 187

Barr Diagnostic Tests, 247; evaluation of, 250; example of, 247-249
Bennett: pupil errors in chemistry, 317; results of Bennett's study, 317-319
Bibliography: arithmetic, 106; commercial subjects, 399; construction of achievement tests, 539; criteria for judging standardized tests, 42; foreign language, 363, 364; general, 7; geography, 281; health and physical education, 420, 421; history and civics, 265-267; industrial arts and vocational subjects, 382-384; intelligence tests, 461-462; interpreting and reporting scores, 66; language, 167; measurement in secondary mathematics, 130; measuring and reporting achievement, 29; miscellaneous, 8, 540; music and art, 300-301; new-type examinations, 508-510; prognosis and special ability tests, 477-480; reading, 232-234; science, 334-335; spelling and handwriting, 193-195
Binet-Simon Tests—Stanford Revision: dependent upon language ability, 450; value of, 449, 450
Biology: diagnostic and remedial work in, 312; tests in, 309-311, 330, 331
Blackstone, E. G., example of test in typing, 388; norms of test, 389; test in typing, 387; use of tests, 387; value of test, 390
Bobbitt, Franklin: objectives in health and physical education, 401, 402
Bookkeeping: Carlson Bookkeeping Tests, 395, 396
Brooks: factors in reading, 199
Brown-Woody: civics test, 256; examples of test, 257, 258
Buckingham-Stevenson: geography tests: content of, 271; excerpt from, 273
Buswell, G. T.: Diagnostic Chart for Individual Difficulties in Arithmetic, 97; excerpts from Chart for Fundamental Processes in Arithmetic, 97-99; diagnosis of Latin reading, 349

Chapman, J. Crosby, 379
Charters, W. W., diagnostic language tests, 142; example of tests, 143
Chemistry: evaluation of Powers' test, 316; examples of test, 314-316; most important objectives, 313; Powers' General Chemistry Test, 314; reliability of Powers' test, 316; value of diagnostic study in, 320
Chemistry tests (*see also name of test*), 314-324; 331
Childs, H. G., revision of Binet-Simon Measuring Scale, 445
Civics: Brown-Woody Test, 256-258; example of Chassell-Upton Scales, 260; how to use Chassell-Upton Scales, 261; important functions of Chassell-Upton Scales, 262; Klapper's three objectives in, 256; Knowlton's definition of, 255; list of tests in, 263; objectives of, 255, 256; purpose of Chassell-Upton Scales, 259
Civics tests (*see also name of test*), 256-263
Clapp Drill Books: content of, 423, 424; exercise from, 425; example of use, 424; good characteristics of, 429; how to diag-

nose with, 427; how to score, 426, 427; how to use, 425; remedial work with, 428; weaknesses of, 429
Commercial Subjects: list of tests, 397-399; measurement of two types in, 386; relation to previous work, 386; status of, 385
Compass Diagnostic Tests in Arithmetic, 82-96; diagnosis with, 91-94; examples from, 89, 90; instructions for giving, 88; objectives measured, 87
Composite: Tressler English Minimum Essentials Test, 154
Composition: objectives measured, 145; ways of measuring, 145
Composition Tests (*see also name of test*), 145-153
Correlation, 60, 61
Courtis, S. A., Standard Research Tests in Music, 295
Courtis Silent Reading Test, No. 2, example and use of, 216-218; interpreting scores made by pupils, 219-220
Criteria for judging standardized tests, 32-42; administration of tests, 38; instructions for giving, 39, 40; interpretation of results, 41; manual of directions, 41; reliability of, 36; scoring the test, 40; validity of, 33
Curtis, F. D.: example of scientific attitude test, 308

Deferrari, R. J., and Foran, T. G.: Latin test, 342
Diagnosis: examples of pupil, 99, 100; in algebra, 119
Diagnostic chart: 93, 118, 139, 149, 154, 178-179, 188, 222, 348, 376
Diagnostic Tests, 82; value of, 414; arithmetic, 97-100
Douglas Standard Diagnostic Tests: for elementary algebra, 114; examples of, 115-117
Dvorak, August: General Science Scales, 307, 308

English (*see Language*).
Examinations: as a basis for marks, 22; deficiencies of, 22; marking of the papers of, 24; new-type compared to traditional, 26; objections to new-type, 27; restricted sampling in, 22; value of traditional, 26; written, 22

Foreign Language: lists of tests in, 357-362; objectives measured, 338; problems of measurement, 339; status of, 336
Forney Test in Map Reading, 278; example of, 278
Fossler, use of algebra tests for diagnosis, 117
Freeman, Frank N., chart for diagnosing faults in handwriting, 188; definition of intelligence, 440; samples of spacing in writing, 190
French: American Council Test, 351

Gates, Arthur I., examples of reading test, 224-226; reliability of reading tests, 202; silent reading test, 224
Gates Primary Reading Tests, 208; criticisms of, 214; remedial work with a 2B class, 212
Gates-Strang Health Knowledge Test: average scores in, 413; directions for giving, 405; examples of content of, 406-413; as example of method of validation, 35; procedure in validation of, 404
General Science, 304; Ruch-Popenoe Test, 304
Geography: diagnosis of pupil difficulties in, 279; essay-type questions, 23; list of tests in, 279; objectives measured, 270; objectives of, 268; practice material in, 438; present-day emphasis, 268; problems of measurement in, 269; standard tests measure, 270; status of geography testing, 270
Geography tests (*see also name of test*), 270-281

Geometry: measurement of, 120; status of, 120
Geometry tests (*see also name of test*), 120-127, 129
German: American Council Test, 351
Glenn and Heck: values of reliable tests in physics, 325
Glenn-Obourn Physics Tests: contents of, 327; value of, 328
Glenn-Welton Chemistry Tests: content of test, 320, 321; examples of test, 322-324
Godsey Diagnostic Latin Comprehension Test, 344; example of, 344, 345
Grammar, Kirby Test, 143; content of, 144; example of, 145; reliability of, 144
Graphing: rules of, 63
Gray, W. S., Oral Reading Tests, 204
Gregory - Hagerty Geography Tests, 277
Group Intelligence Tests, 451; Army Test, Alpha and Beta, 451; list of, 458; uses of, 457-461
Guiler: abilities tested by, 152-154; punctuation abilities, 154; study of thirty-nine college students, 152

Hahn-Lackey Geography Scale, 270; example of, 272; use of, 271
Handwriting: Ayres Scale, 186; diagnosis of samples of, 189-190; diagnostic chart of Ayres Scale, 188; Freeman chart in, 188; quality of line of, 189; measurement of, 186; objective of, 184; rate and quality of handwriting of pupils, 184; sample of Ayres Scale, 187; tests in, 193
Hart: geometry tests, 126
Health: Gates-Strang Test, example of method of validation, 35
Health and Physical Education: needs in development of measurement, 403; objectives of, 400, 401; problems of measurement in, 402; status of, 400; tests in, 420; time allotment, 400; types of measures developed, 404
Health tests (*see also name of test*), 35, 404-413, 420
Height—weight—age: interpretation of, 414; tables by Wood and Baldwin, 415, 416
Hillbrand Sight Singing Test: diagnosis with, 291; how to mark errors—example, 294; how used, 291; types of errors, 291-294
History: diagnosis in, 252; lists of tests in, 263-265; objectives of, 241; place in the curriculum, 240; problems of measurement of, 241; United States, essay-type questions, 22
History tests (*see also name of test*), 242-255, 263-265
Hoke, E. R.: class record for scale, 392; example of scale, 391; Measuring Scale for Knowledge of Gregg Shorthand, 391
Hoke Prognostic Test of Stenographic Ability: content of, 475, 476
Home Economics: content of information test in, 370; examples of Illinois test, 371-375; Illinois Home Economics Association Test, 371; measurement in, 369; need of test in, 370; status of, 369; use of Illinois test, 375-377
Hotz, H. G.: evaluation of algebra test, 114; examples from test in algebra, 113; power test, 112

Indiana Composite Achievement Test: distribution of total scores of, 47; results from spelling test, 177-179; scores from, 46
Industrial Arts: list of tests, 380, 381; need for measurement in, 267; objectives of the teaching of, 366; related to aims of junior high school, 367; status of, 365, 366; time allotment in, 365
Industrial Arts Tests (*see also name of test*), 367-378, 380-382
Instructions for giving test, 39

Intelligence: defined, 440; layman's definition, 441
Intelligence quotients: distribution of, 448; how to find, 448; significance of, 449
Intelligence tests, 29, 440 (*see also name of test*), basis of individual mental test, 444; correlation between various mental tests, 442, 443; low correlation of intelligence scores with achievement scores due to, 442-444; development of, 441; difference between achievement tests and, 441, 442; group tests in, 451-461; influence of school training on scores, 442; measures of the manifestation of capacity in, 441
Interpretation of results, 41
Iowa Physics Tests, 326
Iowa Spelling Scales, 173

John, Lenore: Diagnostic Chart for Individual Difficulties in Arithmetic, 97; examples from chart for fundamental processes in Arithmetic, 97-99
Jones, W. F.: Spelling "Demons," 176

Kepner, P. T.: adverse criticisms of test, 240; Background Tests in Social Science, 236; examples from test, 237-240
Kingsbury, Primary Group Intelligence Scale; instructions and examples from, 455-457
Kirby, F. J., Grammar Test: content of test, 144; example of test, 145; reliability of test, 144
Klapper: three objectives in civics, 256
Kline-Carey Measuring Scale, 296; how to use, 297
Knowlton: definition of civics, 255
Kohs, S. C., definition of intelligence, 440
Kwalwasser-Ruch Test, 287; content of, 288; examples from, 288, 289; uses of, 290

Language: abilities in, 133-135; aims or objectives, 132; in terms of usage, composition, grammar, and literature, 132-161; list of tests, 162-167; measurements deficient in, 135; scientific studies of usage of, 136
Language Tests (*see also name of test*), 136-143; 162-167; 351-362
Latin: need for test in, 350; objectives of the study of, 336-338; remedial work in, 349
Latin tests, 340-350; significant use of, 349
Lennes, N. J., Work, Drill and Test Sheets, 436
Literature: expected outcomes of instruction, 160
Literature tests (*see also name of tests*), 160, 161
Lohr-Latshaw Latin Form Test, 345, 346; example of, 347; types of mistakes made by pupils, 348

MacQuarrie Test for Mechanical Ability, 477
Manual of directions, 41, 538
Marks: administrative use of, 13, 14; application of normal curve to, 17, 19; motive power of, 14; per cent of distribution of, 19, 20, 21; purposes of school, 11, 12; relation to ability, 29; social significance of, 15; source of inaccuracy in, 16; written examination as basis for, 22
Mathematics: measurement in secondary, 109 (*see Algebra and Geometry*); tests in, 127-130
Mathematics Tests (*see also name of test*), 103-106, 127-130, 474
Mean, 48-52
Mean deviation, 53
Measurement: beginnings, 3; in secondary mathematics, 109; principles of, 4-7; results of, 65, 66; sources of inaccuracy in, 16, 17
Measurement in secondary mathematics, 109-130; status and content of, 109
Median, 48-52; to obtain, 48
Miner: status of language, 133; technical grammar, 133

Minnick, J. H., abilities in geometry, 120; examples of geometry test, 121, 122
Mode, 50
Modern European History: diagnostic test in, 251
Modern Language, 351; reliability and norms of tests in, 351
Monroe, W. S., reliability of reading tests, 202; scoring of reading tests, 28; Timed Sentence Spelling Tests, 175
Morgan: Seattle Solid Geometry Test Series, 127
Morton: abilities and skills in arithmetic, 69-70; types of errors in reasoning problems, 101
Mosher Sight Reading Test, content of, 291
Murdock, Katherine: description of analytical scales, 378; description of sewing scale, 378
Music: list of tests in, 298-299; measurement of appreciation, 295; measure of capacity in, 283, 284; notebooks in music appreciation, 439; objectives of music teaching, 282-283; status in public schools, 282
Music tests (*see also name of test*), 284-296, 298-299

National Business Ability Tests, 387
New-Type Examinations: advantages of recall types, 485; characteristics of, 484; classification-example, 505, 506; disadvantages or limitations of recall types, 485; likeness to standard tests, 483; recall types—complete recall—completion, 484-487, 503; rules of importance in making, 498-500; steps in developing, 481-483; turning scores into school marks—example, 500
New York Latin Achievement Test, 340
Norms, 41, 529-534; age, 53, 530-531; example of age and grade, 55; grade, 54, 529-530; mental age, 531; percentile, 531-533; standing of class in relation to, 57; subject and educational age, 54; T-scores, 53, 533-534

Odell, C. W., examples of Odell Scale, 254, 255; scales for rating pupils' answers, 253
Oral reading tests, 204
Orleans-Solomon Latin Prognosis Test: example from, 471; School Habit Rating Scale, 473; use of, 472; what it measures, 470
Osburn, W. J., analysis of the cause of failure in mathematics, 119

Partricks Industrial Arts Test, 367; example of, 367-369
Payne, E. George: rating scale in Health and Accident Prevention, 414
Percentile, 53, 531-533
Physical and Athletic Proficiency: Sargent's test, 417
Physical Fitness: Rogers' index of, 418, 419
Physical Education (*see Health*).
Physics Tests (*see also name of test*), 325-328, 332-334
Pintner, Rudolph: intelligence testing, 440; scoring of intelligence tests, 28, 447-449
Pintner, R., and Patterson, O. G., performance scale, 450
Poetry: Abbott-Trabue Exercises in Judging, 160
Posey-Van Wagenen Geography Scales: description of, 274; examples of, 275, 276; how to use, 276, 277
Powers, S. R., assumptions in study of chemistry, 313; evaluation of test, 316; examples of General Chemistry Test, 314-316; reliability of test, 316
Practice Tests: difficulties in use of, 423; in arithmetic, 423-437; in various fields, 437; purposes of, 422
Pressey Diagnostic Tests in English Composition, 148
Pressey-Richards Tests in American History, 250, 251

Prognosis and Special Ability Tests: based on assumptions, 463-466; content of tests for Junior Clerks, 474; example of, 471; factors in forecasting success, 466; MacQuarrie Test for Mechanical Ability, 477; mechanical ability and aptitude, 476; partial list of tests, 469, 470; purpose of, 463; Stenquist Assembly test, 476; value of, 468
Promotion: standards for, 12, 13

Quartiles: to find, 52

Reading: example of use of a reading test, 207; factors in, 199-201; increase by ages in, 200; increase in scores in, 199; list of tests, 228-232; major objectives of, 197; norms of, 208; objectives measured by tests, 227; oral, 204; problems in measurement of, 202; program of, 197; pupils' errors in, 205; reliability of tests in, 202; remedial work in oral, 207; remedial work in, 213; skills, habits, and abilities in, 198; silent reading tests measure, 208; time allotment of, 197
Reading tests (*see also name of test*), 37, 204, 208, 210-211, 214, 217-222, 223-227, 228-232
Recognition Types of Examination: best answer type, 506; examples of multiple-response, 493, 502-504; identification test-example, 496, 497, 508; matching type-example, 496; objections to true-false type, 488-490; organization type-example, 495, 507; rules in the development and use of, 491; true-false questions, 488, 500, 501; value of best-answers type, 494; value of matching type, 496; value of multiple-response, 494; value of rearrangement test, 497
Record Card, individual, 59, 160, 434
Reliability: how to determine, 535; measure of, 36-37; of reading test, 202; relation to P.E. of raw score, 535-536; significance of, 37
Remedial: measures in arithmetic, 102; work in handwriting, 191; work in oral reading, 207; work in spelling, 182
Reporting results of tests, 65
Rice, J. M.: research in spelling, 2
Rogers, F. R.: physical capacity tests, 417; tests used, 418; validity of tests, 418
Rogers Tests of Mathematical Ability, 474
Rollinson, Ethel A.: diagnostic shorthand tests, 393; examples of shorthand reading test, 394, 395
Ruch, G. M.: experiment with traditional examination, 24, 25; reliability of reading tests, 202
Ruch-Cossman Biology Test, 309; as an example of method of validation, 36; evaluation of, 312; manual of directions, 36; use of, 312
Ruch-Popenoe General Science Test, 304; content of, 305; sample of, 305-307; value of, 307
Russell: Black and Russell, rate and quality of handwriting of pupils, 184, 185

Sangren-Woody Reading Tests, 224, 227; composed of, 227
Sargent, D. A.: attempt to measure physical ability, 417
Scales (*see Tests*).
Schorling-Sanford Test in Geometry: a measurement of, 123; examples of, 123-126; norms for, 126
Science: important aim in teaching of, 309; objectives of the teaching of, 302; problems of measurement in, 303; status of—in public schools, 302
Science Tests (*see name of test*), 304-311, 314-316, 322-325, 326-329, 330-334; objectives measured by, 328-329
Scientific Attitude: Curtis test, 308
Scores: class interval, 46; fre-

quency of, 45; how to use equivalent, 80; interpreting and reporting, 43-67; median, 48; unit of grouping, 46; range of, 45; reporting test, 61; results in single term of mean, median, mode, 48-52; tabulating, 44
Scoring, criteria for, 40
Seashore, C. E.: correlation with achievement (Seashore tests), 286-287; description and use of test, 284-286; tests of musical ability, 284
Seattle: Solid Geometry test, 127
Secondary mathematics: measurement in, 109-131
Shorthand: abilities in, 390
Smith, Martz, and Wright, practice material in geography, 438
Smith, Reeve, and Moss: exercises and tests in algebra, 436; value of tests, 437
Smith and Wright, description of Gregory - Hagerty Geography Tests, 278
Social Science tests, 236-240, 262
Social Studies, reasons for, 236; unification of, 235
Somers, G. T.: Pedagogical Prognosis, 468
Spanish: American Council Test, 351
Spelling: causes of misspelling, 181; Iowa Spelling Scales, 173; Jones Spelling Demons, 176; Monroe Sentence Spelling Tests, 175; objectives measured in, 171; objectives not measured, 170; objectives of, 170; problems of measurement in, 171; remedial work in, 182; results of Indiana Composite Achievement Test, 177-179; Starch measurement of spelling ability, 176; status of, 170; tests in, 192; types of misspelling, 180
Spelling tests (*see name of test also*), 173-179, 192, 193
Stallman, objectives of geography, 268
Standardized tests: criteria for judging, 32-42; faults and abuses of, 27, 28; validity of, 33-35
Stanford Arithmetic Examination, 73; examples from, 76, 77; procedure for giving advanced examination, 75; scoring advanced examination, 77
Stanford Achievement Test: equivalent scores, 79; reliability coefficients of, 74
Stanford Revision of Binet-Simon Measuring Scale, 445; dependent upon language ability, 450; derived score, 447; example of, 446; mental age, 447; value of, 449, 450
Starch, Daniel: tests for measurement of spelling ability, 176
Stenquist, J. L.: assembly test, 476; Mechanical Aptitude Test, 477
Stevenson, P. R., causes of failure in arithmetic reasoning problems, 102
Stone Narrative Reading Tests, 219; diagnostic record chart for, 222; examples and use of, 221
Strayer, G. D.: score card, 3; study of "Age and Grade Census," 2
Strength Index: taken from Rogers tests, 418
Subject Quotient, in arithmetic, 80, 81

Tables: certain rules of construction, 61, 63
Teacher: conditions underlying estimates of pupils' ability by, 21
Terman, L. M., definition of intelligence, 440; revision of Binet-Simon Test, 445
Terman Group Test of Mental Ability, 451; examples from, 451-455
Tests: coaching for, 28; criteria for judging standard, 32-42; faults and abuses of, 27, 28; general type, 71; giving, 43; partial list of prognostic and aptitude tests, 468-470; power test, 528; publicity of results of,

66; rate test, 528; repetition of, 28; scale for rating, 34; validity, 33-35
Tests in Commercial Subjects, 387-399
Thorndike, E. L., 1, 2; influence in measurement, 3; quotation on intelligence from, 440
Thorndike-McCall Reading Scale, 37; description of, 214; reliability of, 215; use of, 216; validity of, 215
Thurstone, L. L., examination in typing, 390
Trabue, M. R.: educational achievement chart, 18; Scales for Measuring Judgment of Orchestral Music, 295
Trade Tests: kinds, 379; method of testing in, 379; uses of, 380
Traditional examination: compared to new-type, 26; marking the papers of, 24, 25; undue labor of, 26; unreliability of, 26; value of, 26
Tressler: English Minimum Essentials Test, 154; examples of test, 156, 157, 158; individual pupil chart, 159, 160; objectives measured, 155
T-score: 53, 533-534
Typing: Blackstone Test, 387; Thurston Examination in, 390

Ullman-Kirby Latin Comprehension Test, 340; content of, 340; evaluation of, 341; examples of, 342; measure of ability to comprehend, 340

Validity: example of method of determining, 35, 36, 515-519; of test, 33-35
Vannest, C. G.: Test in Modern European History, 251
Vocational Subjects, 365
Vocational tests, 381, 382

Wait, W. T.: Seattle Solid Geometry Test Series, 127
White, D. S.: Latin test, 343
Willing, W. H.: chart of results of test, 148; composition test, 146, 147
Wilson, G. M.: adverse criticism of test, 142; diagnostic analysis of results of language error test, 139, 140; examples of use of test, 137, 138; Language Error Test, 136-142; norms of test, 141; objective measured in test, 136
Wood and Baldwin: height-weight-age tables, 415, 416
Woody, Clifford, Language aims, 132
Work-Book (*see Arithmetic Work-Book*).
Work, Drill and Test Sheets, 436
Wright, Wendell W., Indiana Composite Achievement Test, 46-47, 177-179; quotation from, 278

DATE DUE

GAYLORD

PRINTED IN U.S.A.

LB
1131
.S64
Smith, Henry Lest
Tests and
measurements
370.15 Sm5

370.15 Sm5 16389

Smith, H. L. & Wright

Tests & measurements

PIEDMONT COLLEGE LIBRARY
LB1131 .S64
Smith, Henry Le / Tests and measurements,
GE7A